MOLECULAR CHARACTERIZATION OF HYDROLYTIC ENZYMES FROM HYPERTHERMOPHILIC ARCHAEA

Wilfried G.B. Voorhorst

Bacterial Genetics Group
Laboratory of Microbiology
Department of Biomolecular Science
Wageningen Agricultural University

Promotor: dr. W.M. de Vos
hoogleraar in de microbiologie

W.G.B. Voorhorst

Molecular characterization of hydrolytic enzymes from hyperthermophilic archaea

Proefschrift
ter verkrijging van de graad van doctor
op gezag van de rector magnificus
van de Landbouwuniversiteit Wageningen,
dr. C.M. Karssen,
in het openbaar te verdedigen
op vrijdag 24 april 1998
des namiddags te vier uur in de Aula

ISBN 90-5485-796-X

Cover: 3-Dimensional models of the catalytic domain core of the hypermostable serine protease pyrolysin showing aromatic-aromatic interactions (front) and ionic-ionic interactions (back). Drawn by Jack Leunissen (CAOS/CAMM Centre, Nijmegen, the Netherlands) using MOLSCRIPT.

Stellingen

1. De conclusie dat *P.furiosus* vijf intracellulaire proteases produceert kan pas hard gemaakt worden als aangetoond is dat soniceren van een celpellet niet resulteert in het vrijkomen van eiwitten die geassocieerd zijn met of gebonden zijn aan de membraan.

   Blumentals, I.I., Robinson, A.S., and Kelly, R.M. (1990) Characterization of sodium dodecyl sulfate-resistant proteolytic activity in the hyperthermophilic archaebacterium *Pyrococcus furiosus*. Appl. Environm. Microbiol. 56, 1992-1998.

2. Het dirigeren van eiwitten naar het periplasma van *Escherichia coli* kan niet worden gedefinieerd als periplasmatische expressie.

   Makrides, S.C. (1996) Strategies of achieving high-level expression of genes in *Escherichia coli*. Microbiol. Rev. 60, 512-538.

3. Op een sodium dodecyl sulfaat polyacryl amide gel kan geen moleculaire massa bepaald worden van een eiwit dat nog actief is.

   Klingeberg, M. Galunsky, B., Sjoholm, Kasche, V., and Antranikian, G., (1995). Purification and properties of a highly thermostable, sodium dodecyl sulfate-resistant and stereospecific proteinase from the extremely thermophilic archaeon *Thermococcus stetteri* Appl. Environm. Microbiol. 61, 3098-3104.

4. Met de opmerking dat bijna alle niet-cytoplasmatische eiwitten geglycosyleerd zijn wordt zomogelijk het grootste deel van het leven op aarde over het hoofd gezien daar er meer leven op aarde is dan alleen de Eucaryoten.

   Imberty, A., and Pérez, S. (1995). Stereochemistry of the *N*-glycosylation sites in glycoproteins. Prot. Engng. 8, 699-709.

5. Voor de identificatie van plaatsen in een eiwit die betrokken zijn bij thermostabiliteit is meer nodig dan de sequentie van één mogelijk gen.

   Völkl, P., Markiewiz, P., Stetter, K.O., and Miller, J.H. (1994) The sequence of a subtilisin-type protease (aerolysin) from the hyperthermophilic archaeum *Pyrobaculum aerophilum* reveals sites important to thermostability. Prot. Science 3, 1329-1340.

6. Thermofiele enzymen bestaan niet.

7. De omschrijving van een promotie onderwerp is sterk afhankelijk van de onderzoekschool waar het deel van moet uitmaken.

8. De fabrikanten van printers zijn het aan hun stand verplicht de standaardinstellingen van verschillende printers te standaardiseren.

9. Een deel van de in de literatuur beschreven heterogeniteit van enzymatische activiteit voor proteases van hyperthermofielen vindt zijn oorsprong in de gevolgde procedure.

10. Het verwijderen van uitvoegstroken op provinciale wegen heeft een nadelig effect op het milieu en de verkeersveiligheid.

11. In het weekend en op de avonden lijkt in sommige gevallen de instelling van een huisarts het meest weg te hebben van een ik blijf liever thuis arts.

12. Reclameborden langs de snelweg moeten snel weg.

Stellingen behorende bij het proefschrift "Molecular characterization of hydrolytic enzymes from hyperthermophilic archaea".
Wilfried G.B. Voorhorst Wageningen, 24 april 1998

Voorwoord

Door middel van dit voorwoord wil ik iedereen bedanken die er toe hebben bijgedragen dat dit proefschrift tot stand is gekomen. De echte waarde van hun bijdrage is niet te beschrijven. Een aantal wil ik hier met name noemen.

Te beginnen bij mijn promoter Prof. dr. Willem M. de Vos. Willem, ik wil je bedanken voor de enthousiaste begeleiding van het onderzoek, je altijd kritische analyses en het ongelofelijke tempo waarmee je de zoveelste versie van een manuscript door wist te worstelen. Het perfectionisme waarmee je te werk gaat heeft zeker zijn sporen achter gelaten. Rik Eggen, jou wil ik bedanken voor het gestelde vertrouwen en je onmiskenbare aanwezigheid als persoon. De waardevolle wetenschappelijke discussies met jou over de eerste resultaten hebben gedragen tot het ontrafelen van de protease puzzel. Ans Geerling als grote steun en toeverlaat voor heel BacGen, je was de sleutelfiguur die de boel draaiende wist te houden en zorgde voor een prima sfeer. Onmiskenbaar hierbij waren de werkgroep etentjes en de snoeppot, zodat menig student en (gast)medewerker met veel plezier terug kijkt naar de BacGen tijd. Joyce Lebbink, je was herkenbaar aanwezig met je uitbundige karakter, enthousiasme. Tevens wil ik je bedanken voor de hulp bij het bepalen van het goede substraat (blauwe epje). Hauke Smidt, je zorgde voor een rustige basis op de kamer. John van der Oost, de discussies met jou over het hoe en wat binnen het *celB* locus waren soms eindeloos, moge de tijd komen dat alle functies duidelijk zijn. Roland Siezen ben ik zeer erkentelijk voor de samenwerking rond pyrolysine en stetterlysine. Yannick Gueguen, your presence in the lab and your work on the endo-glucanase has been a great pleasure. Gerti Schut voor de uitwisseling van resultaten rond de inductie experimenten met *Pyrococcus*. Vele studenten hebben tijdens de promotie een essentiele bijdrage geleverd aan de voortgang van het onderzoek: Peter Steenbakker, Frank Wagener, Annechien Hengeveld, Mariken Gijsen, Han Zendman, Joost Kolkman, Marcel Dijkgraaf, Angela Warner, Fiona Kaper, Wilfred IJkel, Theo Smits, Bas Lemmens, Esther Poelwijk en Vincent Wittenhorst. Alle mede promovendi, AIO's en OIO's en de ander collega's van de vakgroep Microbiologie wil ik hartelijk bedanken voor de alle hulp die ze mij gegeven hebben en voor de onvergetelijke tijd op de vakgroep.

Mijn ouders, broers en zus wil ik bedanken voor alles wat ze door de jaren heen voor mij gedaan hebben. Jeanet, hoewel woorden niet kunnen beschrijven wat je voor me betekent wil ik je bedanken voor je geduld en steun. Lindsay en Danyon jullie hebben een nieuwe dimensie aan mijn leven toegevoegd, één die ik niet had willen missen.

Pa, we hadden je er ontzettend graag bij willen hebben.

Contents

Chapter 1

General Introduction

In the last few decades microorganisms have been isolated from rather uncommon and hostile locations, such as those with high salt concentrations, an extreme pH, or low or high temperatures. Microorganisms isolated from these environments are referred to as extremophiles (Horikoshi, 1997). The most extensively studied group of these extremophiles are the hyperthermophiles, microorganisms that have an optimum temperature for growth above 80°C (Stetter, 1990). Hyperthermophilic microorganisms appear to be widely spread and have been isolated from hot spots located all around the globe, such as hydrothermal vents, black or white smokers, solfataric fields, hot springs and oil wells. With the isolation of *Pyrolobus fumarii*, the upper temperature of life has been set on 113°C (Blöchl *et al.*, 1997).

Except for two bacterial genera, the *Thermotogales* and *Aquifex*, all hyperthermophiles isolated to date belong to the domain of the *Archaea* (Fig. 1 and Table 1). The *Archaea*, formerly archaebacteria, compose together with the *Bacteria* and *Eucarya* the three domains of life (Fig. 1). Their discovery caused a splitting of the prokaryotic kingdom, recognized in the late 70's based on analysis of 16S rRNA sequences (Woese *et al.*, 1990) (Fig. 1).

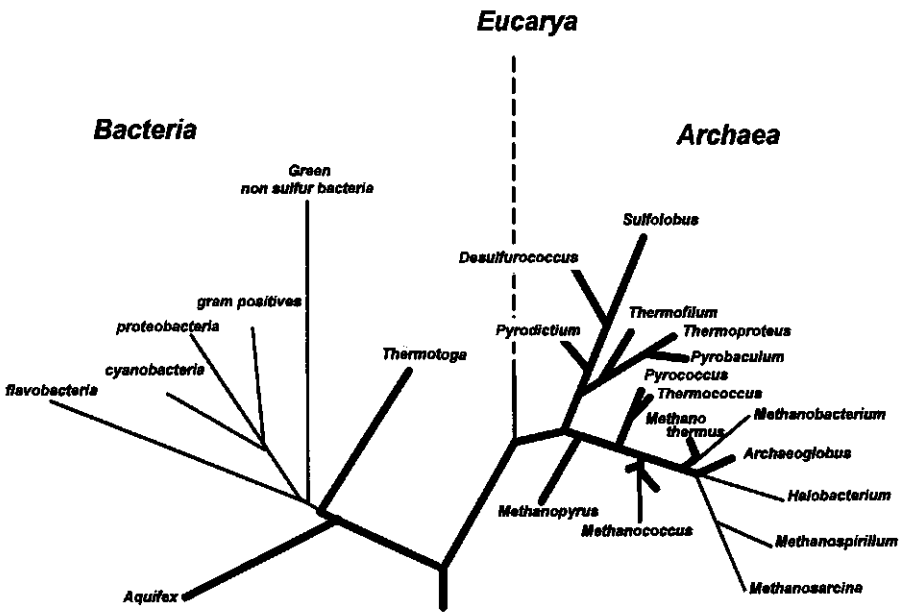


Figure 1. Phylogenetic tree of life based on 16S rRNA sequence analysis. Hyperthermophilic genera are depicted as thick lines (Modified from Woese et al., 1990 and Stetter 1993).

The *Archaea* represent the shortest branches in the phylogenetic tree with hyperthermophiles being the most slowely evolving representatives (Woese *et al.*, 1990). Therefore, it has been suggested that the last common ancestor, the so called 'progenote', could have been a hyperthermophile (Stetter, 1994).

With a few exceptions, all hyperthermophiles grow under strict anaerobic conditions and the most of them depend upon the reduction of elemental sulphur to H_2S for optimal growth (Schönheit and Schäfer, 1995). Except for a limited number of autotrophs, all hyperthermophiles are able to utilize proteins and peptides and several are also able to utilize complex polysaccharides (Table 1) (Kelly and Adams, 1994; Schönheit and Schäfer, 1995). Below an overview is given of (i) hydrolytic enzymes of hyperthermophiles involved in the polymer degradation and the fate of the degradation products in the main metabolic pathways, (ii) the stabilization mechanisms of the enzymes from hyperthermophiles, so-called thermozymes, and (iii) the unique aspects of the molecular biology of *Archaea*. In the discussion of these aspects specific attention will be given to hyperthermophilic archaeon *Pyrococcus furiosus*, since most of the work with hyperthermophiles has been focussed on this versatile microorganism that can grow rapidly, with a doubling time of 37 min, at the normal boiling temperature of water (Fiala and Stetter, 1986). Moreover, *P.furiosus* is able to grow without elemental sulfur and utilize a large range of substrates, including proteins, polysaccharides and pyruvate. Disaccharides, such as cellobiose and maltose, can yield relatively high cell densities, but the monosaccharide glucose can not support growth of *P.furiosus*.

Degradation of polymers in hyperthermophiles

The degradation of polymeric substrates, such as polysaccharides and proteins, is accomplished by the action of various hydrolytic enzymes. Two types of hydrolytic enzymes can be distinguished: (i) hydrolases that act on polymers and cleave internally, so-called endo-acting hydrolases, and (ii) hydrolases that cleaves the terminal residues from either an oligomer or larger polymer, designated exo-acting hydrolases. Below an overview is given of hydrolytic enzymes of hyperthermophiles involved in the degradation of polysaccharides and proteins followed by the main features of their catabolism.

Table 1. Classification and characteristics of hyperthermophilic microorgansims.

Order	Genus	T_{max} (°C)	Polymeric substrates* peptides	polysaccharides
ARCHAEA				
Sulfolobales	*Sulfolobus*	85-87	+	+
	Metallosphaera	80	+	
	Acidianus	75-95	+	
	Stygiolobus	89		
Thermoproteales	*Thermoproteus*	97	+	
	Pyrobaculum	103	+	
	Thermofilum	95	+	+
Desulfurococcales	*Desulfurococcus*	95-97	+	
	Staphylothermus	98	+	
'Pyrolobalus	*Pyrolobus*	113		+
'Pyrodictiales'	*Pyrodictium*	110	+	
	Hyperthermus	108	+	
	Thermodiscus	98	+	+
Themococcales	*Thermococcus*	93-98	+	+
	Pyrococcus	103	+	+
'Archeoglobales'	*Archaeoglobus*	92	+	
Methanobacteriales	*Methanothermus*	97		
Methanococcales	*Methanococcus*	70-91		
'Methanopyrales'	*Methanopyrus*	110		
BACTERIA				+
Thermotogales	*Thermotoga*	84-90	+	
	Thermosipho	77	+	+
	Fervidobacterium	80	+	
'Aquificales'	*Aquifex*	95		

* The capacity to grow on polymeric substrates is listed; 'peptides' indicate complex compounds, e.g. yeast extract, peptone, tryptone, casamino acids and caseine and 'polysaccharides indicate glucose multimers of 2 or more subunits.

Degradation of polysaccharides

The naturally occurring polysaccharides can be divided into two groups depending on the type of linkage between carbohydrate moieties that can be either α-glycosidic or a ß-glycosidic. Enzymes able to hydrolyse these glycosidic bonds are classified as glycosyl hydrolases. Currently, over 60 different families of glycosyl hydrolases have been identified based on amino acid sequence homologies (Henrissat, 1997; Henrissat and Bairoch, 1996). The endo-acting glycosyl hydrolases cleave a polysaccharide chain of

more than 4 subunits, while the exo-acting glycosyl hydrolases remove the terminal glycosidic linkage from usually smaller polysaccharides (less than 5 subunits) (Bauer *et al.*, 1996). In general, the degradation of polysaccharides is initiated by endo-acting glycosyl hydrolases that depolymerize the large substrate resulting in oligosaccharides and monosaccharides. Exo-acting glycosyl hydrolases cleave the oligosaccharides into subunits that subsequently are metabolized. Recently, several reviews have appeared that summarize the properties of glycosyl hydrolases from microorganisms growing above 70°C (Leuschner and Antranikian, 1995; Bauer *et al.*, 1996; Sunna *et al.*, 1997).

To introduce the work presented in this thesis the ß-glycosyl hydrolases of hyperthermophiles will be described in more detail. A relatively large number of ß-glycosyl hydrolases have been characterized from hyperthermophiles. Many bacterial endo-acting and exo-acting glycosyl hydrolases have been identified in *Thermotogales* spp. and these have been reviewed recently (Leuschner and Antranikian, 1995). In contrast, only a few endo-acting ß-glycosyl hydrolases, both endoglucanases and endoxylanases, have been characterized from hyperthermophilic archaea (Table 2). One of them is the endo-ß-1,3-glucanase from *P.furiosus* that is optimally active between 100 and 105°C (See chapter 4; Gueguen *et al.*, 1997). A number of exo-acting ß-glycosyl hydrolases have been described in hyperthermophilic archaea, including one xylosidase and various glycosidases, that all belong to the family 1 of glycosyl hydrolases (Table 2). Notably, the extensively studied *P.furiosus* appears to be well equiped for growth on a range of complex polysaccharides as judged by the presence of various ß-glycosyl hydrolases (Table 2).

Table 2. β-Glycosyl hydrolases that have been characterized in hyperthermophilic archaea.

Type*	Enzyme	Gene	Organism	References
endo-acting	endo-β-1,3-glucanase	*lamA*	*Pyrococcus furiosus*	Gueguen *et al.*, (1997)
	endoxylanase		*Pyrodictium abysii*	Andrade *et al.*, (1996)
exo-acting	β-glucosidase	*celB*	*Pyrococcus furiosus*	Kengen *et al.*, (1993) Voorhorst *et al.*, (1995)
	β-glycosidase	*bglX*	*Pyrococcus furiosus*	Verhees *et al.*, (1997)
	β-mannosidase	*bmn*	*Pyrococcus furiosus*	Kengen *et al*, (1996) Bauer *et al.*, (1996)
	β-glycosidase	*lacS*	*Sulfolobus solfataricus* MT4	Nucci *et al.*, (1993)
	β-glycosidase		*Sulfolobus solfataricus* P2	Grogan (1991)
	β-xylosidase		*Pyrodictium abysii*	Andrade *et al.*, (1996)

* Type of glycosyl hydrolase depending on the site of cleavage: endo-acting, cleavage in a polymer of more than 4 subunits; exo-acting cleavage from the terminal residues, usually of 4 or less subunits.

The pyrococcal ß-glucosidase, CelB, has been characterized in considerable detail and with a half-life value of 85 h at 100°C it is the most thermostable glycosyl hydrolase known to date (Kengen *et al.*, 1993; Kengen and Stams, 1994). The *celB* gene encoding this β-glucosidase was isolated and analysis of the deduced amino acid sequence showed that it belonged to the family 1 of glycosyl hydrolases (Chapter 2; Voorhorst *et al.*, 1995). Functional expression of the *celB* gene in *E.coli* resulted in an enzyme with the same kinetic and stability properties as that purified from *P.furiosus*, which allowed for mutational analysis of the ß-glucosidase to gain insight in structure-function relations (Chapter 2; Voorhorst *et al.*, 1995). Recently, a ß-mannosidase has been characterized from *P.furiosus* (Bauer *et al.*, 1996; Kengen *et al.*, 1996). The *bmn* gene coding for this ß-mannosidase was cloned and the deduced amino acid sequence showed this enzyme also is a member of the glycosyl hydrolase family 1, with 46.5 % identity to the ß-glucosidase of *P.furiosus* (Bauer *et al.*, 1996).

Like the pyrococcal ß-glucosidase CelB, the homolog from *Sulfolobus solfataricus*, the ß-glycosidase LacS has been extensively studied. The *lacS* gene has been expressed in different hosts, including mammalian cells, and yeast, and used as reporter gene (Cannio *et al.*, 1994; Moracci *et al.*, 1992). The substrate specificity of the β-glycosidase as well as effects of temperature and SDS have been studied (Nucci *et al.*, 1993; Nucci, *et al.*, 1995). Mutational analysis showed that two glutamate residues are essential for catalysis (Moracci *et al.*, 1996). The β-glycosidase from *S.solfataricus* has been used in transglycosylation, however, it was only able to accept secondary alcohols as non-sugar aglycons in the synthesis reaction, whereas the pyrococcal β-glucosidase was also able to use tertiary alcohols (Fisher *et al.*, 1996; Trincone *et al.*, 1994; Trincone and Pagnotta, 1995). Recently, the crystal structure of the *S.solfataricus* LacS has been elucidated and showed a large number of ion-pairs that are involved in networks and a relatively large number of solvent molecules that are buried in hydrophilic cavities (Aguilar *et al.*, 1997).

The metabolism of glucose, an endproduct of glucose polymer hydrolysis, has been studied in hyperthermophiles belonging to the bacterial genus *Thermotoga* and in different genera of *Archaea* such as *Pyrococcus, Sulfolobus, Desulfurococcus* and *Thermoproteus*. The hyperthermophilic bacterium *Thermotoga maritima* appeared to use the classical Embden-Meyerhof pathway (Schönheit and Schäfer, 1995; Selig *et al.*, 1997). In contrast, in the hyperthermophilic archaea modifications have been identified in the Embden-Meyerhof pathway and in the Entner-Douderoff pathway [for reviews see Schönheit and Schäfer, 1995; Kengen *et al.*, 1996; Selig *et al.*, 1997].

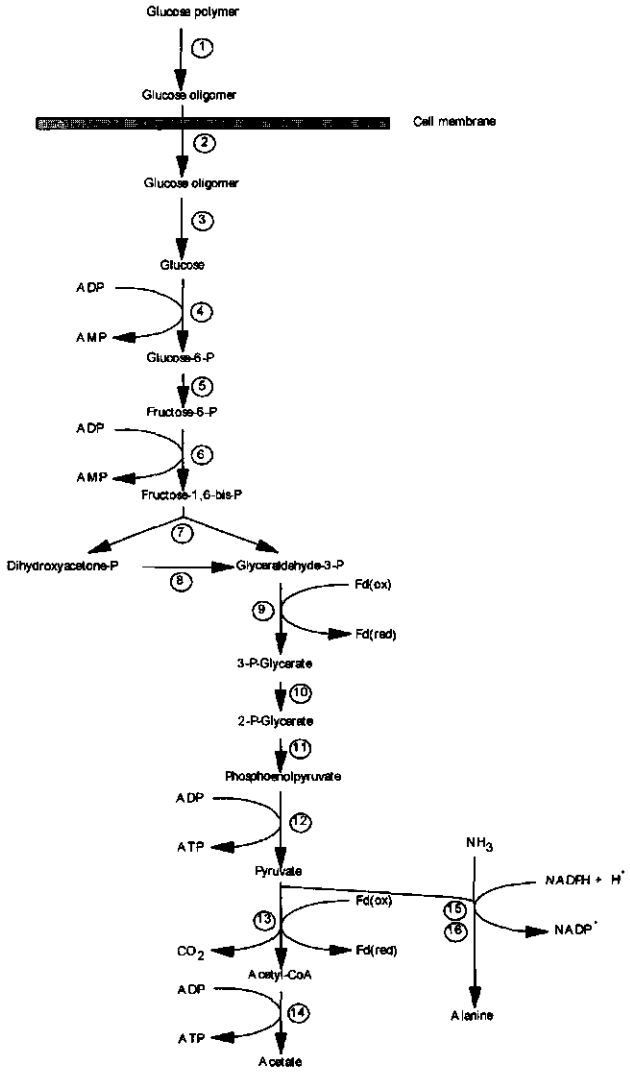


Figure 2 Proposed pathway of glucose fermentation in *P.furiosus*. Glucose polymer, consisting of more than four glucose subunits; glucose oligomer, consisting of 2 to 4 glucose subunits.1, endo-acting glycosyl hydrolase; 2, unknown transport system; 3, endo-acting glycosyl hydrolase; 4, ADP-dependent glucokinase; 5, phosphoglucoisomerase; 6, ADP-dependent phosphofructokinase; 7, fructose-1,6-biphosphate aldolase; 8, triosephosphate isomerase; 9, glyceraldehyde-3-phosphate:ferredoxin oxidoreductase; 10, phosphoglycerate mutase; 11, enolase; 12, pyruvate kinase; 13, pyruvate:ferredoxin oxidoreductase; 14, acetyl-CoAsynthetase (ADP-forming); 15, glutamate dehydrogenase; 16, alanine aminotransferase; Fd, ferredoxin. (Modified from Kengen *et al.*,1996).

Modifications observed in the Embden-Meyerhof pathway used by *P.furiosus* and *Thermococcus* spp. include the presence of two ADP-dependent kinases, glucokinase and phosphofructokinase (Fig. 2) (Kengen *et al.*, 1994; Kengen *et al.*, 1996). Additionally, a glyceraldehyde 3-phosphate ferredoxin oxidoreductase has been identified in *P.furiosus* which converts glyceraldehyde-3-phosphate to 3-phosphoglycerate, without substrate phosphorylation (Fig. 2) (Mukund and Adams, 1995; van der Oost *et al.*, 1997). This contrasts with the classical glycolysis that uses glyceraldehyde-3-phosphate dehydrogenase. Low levels of this glyceraldehyde-3-phosphate dehydrogenase were also identified in *P.furiosus*, but the enzyme is thought to function in gluconeogenesis (Schäfer and Schönheit, 1993; van der Oost *et al.*, 1997). A partially non-phosphorylated Entner-Douderoff pathway has been established in *Sulfolobus* spp. (De Rosa *et al.*, 1984) and in *Thermoplasma acidophilum* (Budgen and Danson, 1986; Danson, 1988; Danson, 1989). *Thermoproteus tenax*, one of the deep branches of the *Archaea,* was found to contain the Embden-Meyerhof pathway with an ATP-dependent hexokinase and a PP_i-dependent phosphofructokinase (Siebers and Hensel, 1993). However, intermediates of the non-phosphorylated Entner-Douderoff pathway have been also been identified this archaeon (Hensel *et al.*, 1987). In all *Archaea* and hyperthermophilic bacteria, analyzed so far, the conversion from pyruvate to acetate involves an oxidative decarboxylation reaction by pyruvate:ferredoxin oxidoreductase to yield acetyl-CoA (Schönheit and Schäfer, 1995). The conversion of acetyl-CoA to acetate appears to be a catalyzed by an ADP-forming acetyl-CoA synthetase in all acetate-forming archaea (Schäfer *et al.*, 1993).

Degradation of proteins and peptides

Proteases are the key enzymes in the conversion of proteinaceous substrates to peptides and amino acids. They have been classified based on their active center and the following classes can be distinguished: serine, cysteine, aspartic, or metallo proteases (Bairoch, 1997). During catalysis serine and cysteine proteases form a covalent acyl-enzyme intermediate, while metallo and aspartic proteases activate a water molecule that subsequently attacks the substrate. Serine proteases have an active centre composed of a catalytic triade of three residues, aspartate, histidine and a serine, that form a charge relay system, and an asparagine that stabilizes the oxyanion generated in the transition state (Carter and Wells, 1990). Proteases from this class are sensitive to inhibition by

phenylmethylsulphonyl fluoride and diisopropylfluorophosphate. The serine proteases can be divided into many subgroups, with the two major groups being the subtilisin-like and trypsin-like serine proteases (Siezen *et al.*, 1991). The subtilisin-like serine proteases have been extensively studied in many other organisms with respect to structure-function and structure-stability relations [for review (Siezen *et al.*, 1991)]. Cysteine proteases contain a cysteine and a histidine residue within the active site. Aspartic proteases (also referred to as acid proteases) contain two active-site aspartic acid residues in close proximity, one of which is ionized at low pH. Metallo proteases contain divalent cations, mainly zinc, that are chelated to two or more histidine residues, while a glutamate acts as a catalytic base. Thermostable proteases have recently been reviewed (Daniel *et al.*, 1995).

Various proteases have been identified in hyperthermophilic genera and many of those were classified as serine proteases (Table 3). The first archaeal serine proteases was isolated from cell-free supernatant of the hyperthermophile *Desulfurococcus* strain $Tok_{12}S_1$, and designated archaelysin (Cowan *et al.*, 1987). More recently, from the related strain, *Desulfurococcus* strain SY, a cell-associated serine protease was characterized that showed a half-life of 4.3 hours at 95°C (Hanzawa *et al.*, 1996). Various proteases have been characterized from the acidophilic *Sulfolobus solfataricus* strains, including two serine proteases (Fusi *et al.*, 1991). The properties of extracellular proteases from a number of *Thermococcus* species have been analyzed and found to be serine proteases composed of multiple activity bands (Klingeberg *et al.*, 1991). From one of these, *T.stetteri*, the extracellularly located serine protease has been characterized. This 68 kDa protease was highly stable and resistant to chemical denaturation as illustrated by a half-life of 2.5 h at 100°C and retention of 70% of its activity in the presence of 1% SDS (Klingeberg *et al.*, 1995).

A globular serine protease has found to be associated with the stalk of a filiform glycoprotein complex, termed tetrabrachion, at the surface of the *Staphylothermus marinus* (Mayr *et al.*, 1996). This serine protease, designated STABLE, was found to be extremely thermostable especially if bound to the tetrabrachion where residual activity could still be detected after 10 min incubation at 135°C (Mayr *et al.*, 1996). The gene encoding STABLE has been isolated and the deduced sequence showed highest homology with subtilisin-like serine proteases, although a large insert was found within the catalytic domain and a relatively large C-terminal extension was observed (Mayr *et al.*, 1996). Another gene encoding a subtilisin-like serine protease, designated aerolysin, has been cloned from *Pyrobaculum aerophilum* and its deduced sequence was modelled

Table 3. Extra- and intracellular proteases of hyperthermophilic microorganisms.

Organism	Enzyme	Protease class	Gene	Classification and features*	References
ARCHAEA					
Pyrococcus furiosus	pyrolysin	serine	+	pyrolysin, G	Eggen *et al* (1990), Voorhorst *et al.*, (1996)
	PfpI(S66)	new type			Halio *et al.*, (1996)
	prolyl endopeptidase	endopeptidase	+		Harwood *et al.*, (1997),Robinson *et al.*, (1995)
Pyrococcus sp. KOD1		thiol			Fujiwara *et al.*, (1996)
Pyrodictium abyssi		serine	+	subtilisin	Voorhorst (unpublished data)
Pyrobaculum aerophilum	aerolysin	serine	+	subtilisin, G	Volkl *et al.*, (1994)
			+	serine protease⁺	Fitz-Gibbon *et al.*, (1997)
			+	subtilisin⁺	
			+	dipeptidase⁺	
Desulfurococcus mucosus	archaelysin	serine			Cowan *et al.*, (1990)
Desulfurococcus strain SY		serine			Hanzawa *et al.*, (1996)
Thermococcus stetteri		serine			Klingeberg *et al.*, (1991, 1992, 1994)
	stetterlysin	serine	+	pyrolysin, G	Voorhorst *et al.*, (1997)
Thermococcus celer		serine			Klingeberg *et al.*, (1991, 1992, 1994)
Thermococcus litoralis		serine			Klingeberg *et al.*, (1991, 1992, 1994)
Thermococcus AN1		serine			Klingeberg *et al.*, (1991, 1992, 1994)
Sulfolobus solfataricus					Burlini *et al.*, (1992)
Sulfolobus acidocaldarius	thermopsin	new acid protease	+		Ling and Tang (1990),Fusek *et al.*, (1990)
Staphylothermus marinus		serine			Klingeberg *et al.*, (1991, 1992, 1994)
	STABLE	serine	+	pyrolysin, G	Mayr *et al.*, (1996)
BACTERIA					
Thermobacteroides proteolyticus					Klingeberg *et al.*, 1991
Fervidobacterium pennavorans		serine			Klingeberg *et al.*, (1992)
			+	subtilisin	Voorhorst (unpublished data)

* Classification into known families of subtilisin-like serine proteases as defined by Siezen and Leunissen (1997) are indicated; ⁺ derived from a genome sequencing project; G, indicates the presence of putative *N*-glycosylation sites.

on known structures of subtilisin-type proteases (Völkl *et al.*, 1995).

Multiple proteolytic activities have been observed in *P.furiosus*, consisting of 5 to 13 activity bands with different molecular weights on SDS-PAGE (Blumentals *et al.*, 1990; Eggen *et al.*, 1990; Connaris *et al.*, 1991). Two of these, S66 and S102, were found to be SDS-resistant, but showed no immunological cross reaction (Blumentals *et al.*, 1990). More recently, it was found that the S66 protease was a homomultimer consisting of 18.8-kDa subunits, designated PfpI (Halio *et al.*, 1996). Only the tetramer and some higher aggregation forms showed proteolytic activity, but not the monomeric subunits The *pfpI* gene was cloned and functionally expressed in *E.coli*, although the specific activity of the recombinant enzyme was lower than observed with the enzyme purified from *P.furiosus*. The deduced amino sequence of PfpI showed no conserved protease motifs nor significant homology with any other known protease (Halio *et al.*, 1996).

The cell-envelope associated protease activity of *P.furiosus* was designated pyrolysin and was found to be a highly stable serine protease activity with a half-life of 20 min at 105°C (Eggen *et al.*, 1990). The multiple proteolytic bands appeared to be processed during autoincubation from high molecular weight bands to give a final 65 kDa activity band (Eggen *et al.*, 1990). Recently, pyrolysin was purified from membrane fractions of *P.furious* resulting in a high and a low molecular weight fraction pyrolysin, the latter form being a processing product of the high molecular weight fraction (Chapter 6; Voorhorst *et al.*, 1996). Both forms of pyrolysin were glycosylated and active towards a broad range of substrates. The *pls* gene, encoding pyrolysin, was cloned via reversed genetics and its deduced sequence analysed. It was shown that pyrolysin is a subtilisin-like serine protease, based the presence of the high homology with representatives from this class of proteases (Chapter 6; Voorhorst *et al.*, 1996).

Other types of proteases identified from hyperthermophiles are thermopsin from *S.acidocaldarius*, a thiolprotease from *Pyrococcus* sp. KOD1 (Morikawa *et al.*, 1994; Fujiwara *et al.*, 1996) and a prolylpeptidase (PEPase) from *P.furiosus* (Robinson *et al.*, 1995; Harwood *et al.*, 1997). Thermopsin, a new type of acidic protease, has been extensively characterized and showed optimal activity at 75°C and pH 2 (Fusek *et al.*, 1990). The gene encoding thermopsin was cloned and its deduced sequence predicted the presence of a large number of *N*-glycosylation sites (Lin and Tang, 1990). From hyperthermophilic bacteria only a few proteases have been identified, including an extracellular serine protease from *Thermobacteroides proteolyticus* (Klingeberg *et al.*,

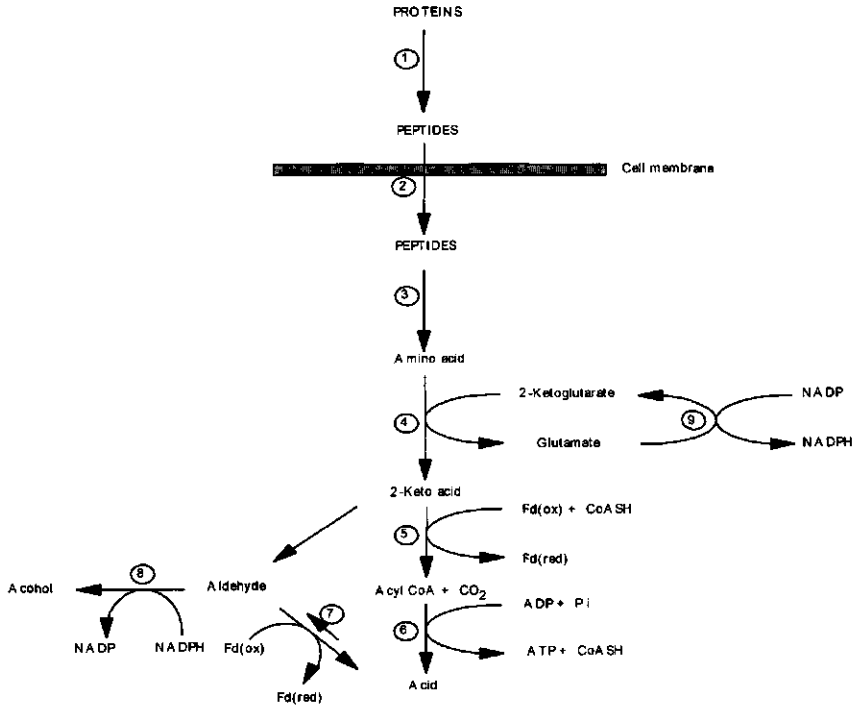


Figure 3. Proposed pathway for peptide fermentation by S°-dependent hyperthermophiles. 1, extracellular protease; 2, unknown peptide transport system; 3, amino and other peptidases; 4, transaminase; 5, 2-keto acid ferredoxin oxidoreductases (POR, IOR, KGOR, VOR); 6 acetyl-CoA synthetase; 7, aldehyde ferredoxin oxidoreductase; 8 alcohol dehydrogenase; 9 glutamte dehydrogenase CoASH, coenzyme A; Fd, Ferredoxin; (Modified from Mai and Adams, 1996).

1991) and a serine protease that was active toward keratine and other insoluble proteins from *Fervidobacterium pennavorans* (Klingeberg *et al.*, 1992) (Table 3).

Oligopeptidases and dipeptidases, identified in numberous bacteria, have not been characterized in hyperthermophiles, although their presence has been proposed based on sequence homologies (Fitz-Gibbon *et al.*, 1997; Voorhorst, unpublished data).

Only little is known about the intracellular conversion of proteinaceous substrates in hyperthermophiles. Most information has been derived from *P.furiosus* and *Thermococcus litoralis* and a metabolic pathway for amino acids has been proposed in these archaea with the first step being an aminotransferase reaction to yield 2-keto acids (Fig. 3) (Kelly and Adams, 1994). Alternatively, a decarboxylation reaction may occur resulting in an aldehyde (Adams and Kletzin, 1996). Subsequently, the 2-keto

acids are converted by specific ferredoxin oxidoreductases in a CoASH-dependent reaction to form the corresponding activated keto-acid. As primary substrates they may use indole pyruvate (IOR; Mai and Adams, 1994), 2-ketoisovalerate (VOR; Heider *et al.*, 1996) or 2-ketoglutarate (KGOR; Mai and Adams, 1996). Two reversible and ADP-dependent acetyl Coenzyme A synthetase have been proposed to function in *P.furiosus* in the conversion of these keto-acids to yield the corresponding acids (Fig. 3) (Mai and Adams, 1996).

Thermostabilization of proteins from hyperthermophiles

A major reason for the increasing interest in hyperthermophiles is the extreme stability of their proteins and enzymes. Because of their potential in a variety of applications, the latter have also been referred to as thermozymes (Adams, 1993; Vieille and Zeikus, 1996). In addition to their extreme thermostability, thermozymes showed an increased resistance to chemical denaturation and proteolytic degradation compared to their less stable counterparts (Fontana, 1991; Vieille and Zeikus, 1996). The primary stability of proteins is encoded by their amino acid sequence and involves a large number of attracking and repulsing forces. In addition, this intrinsic stability may be further improved by external factors, such as salt or metal-ion-binding [reviewed by Gupta, 1991 and Gray, 1993] and the compatible solutes such as mannosylglycerate and di-*myo*-inositol-phosphate present in *P.furiosus* (Martins and Santos, 1995). Different approaches have been used to identify factors that may explain the relatively high intrinsic stability of proteins from hyperthermophilic origin. The most simple approach is comparison on the amino acid sequences and the overall amino acid compositions of enzymes with different thermostability. These comparisons have indicated a decrease in the chain flexibility and content of chemically labile residues (Asn, Cys) with increasing temperature (Zwickl *et al.*, 1990; Eggen *et al.*, 1994; Vieille and Zeikus, 1996). Additionally, an increased average hydrophobicity and frequency of aromatic residues were suggested to be correlated with higher thermostability (Zwickl *et al.*, 1990; Hess, 1995: for recent review see Vieille and Zeikus, 1996). Sequence alignments of homologous proteins with different thermostability has led to the identification of structural and functional features, such as insertions, deletions, that may contribute to stability. and conservation of (active-site) amino acid residues among homologous enzymes (Bouyoub *et al.*, 1996).

For a more detailed identification of structural features related to the thermostability of enzymes, a comparative structural approach has been used, that include the comparison of three-dimensional structures (for recent reviews see Querol *et al.*, 1996; Vieille and Zeikus, 1996). In the cases that these structures were not available, molecular modelling served as alternative to obtain insight in the structural features related to protein stabilization. This approach resulted in identification a variety of mechanisms by which proteins from hyperthermophiles could be stabilized (Table 4). These mechanisms resemble those identified for mesophilic model proteins (Fersht and Serrano, 1993) and it seems that each thermostable protein is stabilized by a unique combination of different mechanisms resulting in a more rigid structure (Table 4) (Querol *et al.*, 1996; Vieille and Zeikus, 1996). Analysis of the structures identified so far for proteins from hyperthermophiles, suggests that an increased number of ion-pairs, that may be arranged in extensive networks, may be an important factor contributing to the observed extreme stability. However, prospective, protein-engineering based studies are still scarce (Tomschy *et al.*, 1994; Lebbink *et al.*, 1995).

Table 4. Proteins from hyperthermophiles that have been analysed on structural level.

Protein	Organism	Structure*	Stabilization	References
Aldehyde oxidoreductase	*Pyrococcus furiosus*	C, 2.3 Å	- small solvent-exposed surface area - large number of ionpairs and buried atoms	Chan *et al.*, 1995
Glutamate dehydrogenase	*Pyrococcus furiosus*	C, 2.2 Å	- intrasubunit ion pair up to networks networks - decrease intrasubunit cavities by ionic interactions	Yip *et al.*, 1996
	Thermococcus litoralis	C	- ion pair networks	Rice *et al.*, (1996)
	Thermotoga maritima	C, 3.0 Å	- intrasubunit ion pair up to networks - intersubunit ion pairs reduced and increase of hydrophobic interactions compaired to GDH *P.furiosus* - decrease intrasubunit cavities by hydrophobic interactions	Knapp *et al.*, (1997)
β-Glycosidase	*Sulfolobus solfataricus*	C, 2.6 Å	- surface ionpairs involve in networks that cross-link separate structures - large number of solvent molecules buried in hydrophilic cavities in the protein core	Aguilar *et al.*, (1997)
Indole-3-glycerol phosphate synthase	*Sulfolobus solfataricus*	C, 2.0 Å	- ion pairs, - α-helices - strenghtening loops and termini	Hennig *et al.*, 1995
Holo-glyceraldehyde 3-phosphate dehydrogenase	*Thermotoga maritima*	C, 2.5 Å	- ion pair formation	Korndörfer *et al.*, 1995
TATA-binding protein		C, 2.2 Å	- S-S bridge - increased number of surface ionpairs - more compact protein packing	DeDecker *et al.*, 1996
Rubredoxin	*Pyrococcus furiosus*	N	- minimal structural variations at room temperature - electrostatic interactions	Blake *et al.*, 1992 Cavagnero *et al.*, 1995
Ferredoxin	*Pyrococcus furiosus*	N	- extended structural elements - tighter turns	Teng *et al.*, 1994
Pyrolysin	*Pyrococcus furiosus*	M	- ion pair formation up to networks - aromatic-aromatic interactions - helix dipole stabilization	Voorhorst *et al.* 1997
Stetterlysin	*Thermococcus stetteri*	M	- ion pair formation up to networks - aromatic-aromatic interactions - helix dipole stabilization	Voorhorst *et al.*, 1997
Aerolysin	*Pyrobaculum aerophilum*	M	- alanine transitions - helix dipole stabilization by aspartate	Völkl *et al.*, 1994
Citrate synthase	*Pyrococcus furiosus*	M	- reduced length of the loops - stabilization α-helix - increased negative charge on the surface - increased frequency of aromatic residues	Muir *et al.*, 1995

* type of three-dimensional structure: C, structure based on crystal structures with resolution of the crystals; N, structure based on NMR data; M, structure based on molecular modelling

Molecular biology of hyperthermophiles

The molecular biological analysis of hyperthermophiles has mainly been focussed on the archaeal representatives due to their unique phylogenetic position. Archaeal genomes seem to have components that show resemblance to that of either *Bacteria* or *Eucarya*. On the one hand, *Archaea* resemble *Bacteria* with respect to components of the translation machinery, genome size and clustering of genes (Table 5). On the other hand, the transcription systems of *Archaea* show more resemblance to that of *Eucarya* (Table 5) (Baumann *et al.*, 1995; Langer *et al.*, 1995; Thomm, 1996; van der Oost, *et al.*, 1997). The archaeal RNA polymerase is structurally related to the RNA polymerases II and III of *Eucarya* and shows a high degree of complexity like its eucaryal counterparts (Langer *et al.*, 1995). The archaeal promoter, closely resembles the eucaryal RNA polymerase II promoters, with an eucaryal TATA-box located 25 nucleotides upstream from the transcription initiation site. The archaeal transcription factors, such as TATA-binding protein (TBP) and RNA polymerase II transcription factor B (TFIIB), are highly homologous to those of *Eucarya* and are not found in *Bacteria*. Moreover, eucaryal TBPs were found to be functionally exchangeable with their archaeal homologs in archaeal *in vitro* transcription systems (Hethke and Thomm, 1997; Wettach *et al.*, 1995). The existence of eucaryal-like gene-specific regulators has been predicted for *Archaea* based on the high conservation of TBP and TFIIB that in *Eucarya* directly interact with regulators (Baumann *et al.*, 1995). However, this possibility is not compatible with the recent structural analysis of the TATA-binding protein of *Pyrococcus woesei*, which showed that the site for regulatory interaction is different from the eucaryal homologous, whereas the DNA-binding site is highly homologous (Dedecker *et al.*, 1996). In two instances, bacterial-like negative regulator binding sites have been identified in *Archaea* composed of a palindromic structure (Ken and Hackett, 1991; Stolt and Zillig, 1992; Cohen-Kupiec *et al.*, 1997).

The size of the archaeal genomes is comparable to that of *Bacteria* and the clustering of genes into operon-like organizations has been reported (Baumann *et al.*, 1995; Tutino *et al.*, 1993; Pedroni *et al.*, 1995; van der Oost, *et al.*, 1997; Voorhorst *et al.*, 1997: Chapter 4). Bacterial genes located in an operon are in general functionally related. However, for some *Archaea* the functional relation of the genes in an operon remained unclear (chapter 4). A similar observation has been made following complete genome sequence analysis in the hyperthermophilic bacterium *Aquifex aelicus* where operons have been observed with a mixture of unrelated genes (Fox, 1997).

Table 5. Some discriminating molecular features of the three domains of life.

	Bacteria	*Archaea*	*Eucarya*
Genome	single, circular, haploid	single, circular, haploid	multiple, lineair, diploid
Plasmids	+	+	-
polycistronic transcripts	+	+	-
RNA polymerase	one class ($\beta\beta'\alpha_2$)	one class (8-10 subunits)	three classes, complex composition
promoter signatures	-35/-10	TATA-box (-25)	TATA-box (-25)
Transcription factors	σ	TBP TFIIB	TBP TFIIB
Translation initiation	SD + AUG	SD + AUG	AUG

SD, Shine Dalgarno sequence; σ, sigma factor

It has been postulated that control of archaeal genes resembles that found in *Bacteria* (Zillig *et al.*, 1993). Bacterial-like transcriptional regulators have been identified on the sequence level such as LacI homolog in *P.aerophilum* (Fitz-Gibbon *et al.*, 1997) and homologs of the *E.coli* general activator Lrp have been identified in *Pyrococcus, Sulfolobus* and *Methanococcus* (Kyrpides and Ouzounis, 1995: Lebbink *et al.*, 1997; Sensen *et al.*, 1996; Bult *et al.*, 1996), but their functionality remains to be proven.

Aims and outline of the thesis

The aims of the research presented in this thesis were (i) to characterize hydrolases from hyperthermophilic archaea and to analyse the possibilities to functionally produce these hyperthermostable enzymes in heterologous expression systems, (ii) to gain insight in the structure-function and structure-stability relation of these enzymes, and (iii) to study regulation of gene expression in hyperthermophilic archaea. The research was focussed on two groups of hydrolytic enzymes, glycosyl hydrolases and proteases, since they are key enzymes in degradation of the main substrates for hyperthermophiles, polysaccharides and proteins, respectively. Moreover, these clasess of hydrolytic enzymes have been extensively studied over a large temperature range from psychrophilic, mesophilic and thermophilic origin and therefore can be used in comparative studies to gain insight in features related to protein stabilization.

In this study, the hyperthermophilic archaeon *Pyrococcus furious* was used as model organism, since it is able to grow on a wide range of substrates including proteins, peptides, and carbohydrates. Meanwhile, *P.furiosus* is being developed into a model system for hyperthermophiles.

In chapter 1 an introduction is given into various aspects of hyperthermophiles, including an overview of hydrolytic enzymes involved in the degradation of polymers and the catabolic pathways of the monomers, the approaches used to analyse protein stability and the present state of the art, and some characteristics of the molecular biology of *Archaea*.

During growth on cellobiose *P.furiosus* showed high levels of an intracellular ß-glucosidase, able to hydrolyse the ß-1,4-glycosidic bond of this disaccharide (Kengen *et al.*, 1993). In chapter 2 it is described that the *P.furiosus celB* gene encoding this ß-glucosidase was isolated, characterized and found to belong to the glycosyl hydrolase family 1. The *celB* gene was functionally overexpressed in the mesophilic host *Escherichia coli* resulting in a ß-glucosidase with stability and kinetic properties that were indistinguishable from that purified from *P.furiosus*. Mutational analysis, to gain insight in structure-function relation, indicated that the hyperthermostable representative of the glycosyl hydrolase family 1 uses the same catalytic mechanism as identified for a representative from mesophilic origin. The analysis of the genomic region upstream of the *celB* gene showed the presence of a gene cluster that comprised two alcohol dehydrogenases, and an additional glycosyl hydrolase with endo-ß-1,3-glycosyl

hydrolase activity. The organization, expression and function of this gene cluster is studied in the following chapters (Chapter 3-5).

Chapter 3 describes the overproduction and characterization of the two alcohol dehydrogenases, a short-chain and an iron-containing alcohol. Chapter 4 deals with the expression in *E.coli* of the endo-ß-1,4-glucanase encoding gene, *lamA*, of the *celB*-locus and characterization of the produced hydrolase. The endo-ß-1,3-glucanase hydrolyses ß-1,3 and 1,4-glycosyl bonds, generating oligosaccharides that subsequently may be converted to monosaccharides by the ß-glucosidase CelB. Mutational analysis was performed to analyse the substrate specificity of the enzyme. The gene cluster containing the *adhA-adhB-lamA* genes was found to be a operon, that was designated the *lamA* operon. The *lamA* transcript was found to be co-regulated with the divergently orientated *celB* gene by ß-linked glucose polymers as is described in Chapter 5.

Chapters 6 and 7 are concerned with the second group of hydrolytic enzymes studied, the extracellular subtilisin-like serine proteases. Chapter 6 describes the purification and characterization of pyrolysin, an abundant extracellular protease in *P.furiosus* when growing on proteinaceous substrates. Pyrolysin, one of the most thermostable serine proteases known to date, was found to be N-glycosylated and autocatalytically processed at its C-terminus. Via reversed genetics the gene was isolated and its sequence showed to be a subtilisin-like serine protease (Chapter 6). A pyrolysin-like encoding gene was also isolated from *Thermococcus stetteri*. The deduced product, designated stetterlysin, showed high homology with pyrolysin and with subtilases that have a known three-dimensional structure allowing for homology modelling. Comparisons have been performed of the predicted three-dimensional models of the catalytic domain of stetterlysin and pyrolysin with the crystal structure of subtilases from mesophilic and thermophilic origin and a model structure for a subtilase from psychrophilic origin. This resulted in the identification of features that could be related to protein thermostabilization (Chapter 7). Modelling of substrates into the predicted structure revealed the possible enzyme-substrate interactions and suggest that pyrolysin is able to remove its own pro-peptide and thus is autocatalytically activated. A summary and concluding remarks are given in Chapter 8.

References

Adams, M.W.W. (1993) Enzymes and proteins from organisms that grow near and above 100°C. *Ann. Rev. Microbiol.* **47**: 627-658.

Adams, M.W.W., and Kletzin, A. (1996) Oxidoreductase-type enzymes and redox proteins involved in fermentative metabolisms of hyperthermophilic Archaea. *Adv. Prot. Chem.* **48**: 101-180.

Aguilar, C.F., Sanderson, I., Moracci, M., Ciaramella, M., Nucci, R., Rossi, M., and Pearl, L.H. (1997) Crystal structure of the β-glycosidase from the hyperthermophilic archaeon *Sulfolobus solfataricus*: Resilience as a key factor in thermostability. *J. Mol. Microbiol.* (In press).

Andrade, C.M., Morana, A., de Rosa, M., and Antranikian, G. (1996) Production and characterization of amylolytic and xylanolytic enzymes form the hyperthermophilic archaeon *Pyrodictium abyssi*. First International Congres on Extremophiles, Estoril, Portugal, p 98.

Bauer, M.W., Bylina, E.J., Swanson, R.V., and Kelly, R.M. (1996) Comparison of a β-glucosidase and a β-mannosidase from the hyperthermophilic archaeon *Pyrococcus furiosus*. Purification, characterization, gene cloning, and sequence analysis. *J. Biol. Chem.* 271: 23749-23755.

Bauer, M.W., Halio, S.B., and Kelly, R.M. (1996) Proteases and glycosyl hydrolases from hyperthermophilic microorganisms. In *Enzymes and Proteins from Hyperthermophilic Microorganisms*. NY, Academic Press. pp. 271-310.

Baumann, P., Qureshi, S.A., and Jackson, S.P. (1995) Transcription: new insights from studies on *Archaea*. *Trends Genet.* **11**: 279-283.

Blake, P.R., Park, J.-B., Zhou, Z.H., Hare, D.R., Adams, M.W.W., and Summers, M.F. (1992) Solution-state structure by NMR of zinc-substituted rubredoxin from the marine hyperthermophilic archaebacterium *Pyrococcus furiosus*. *Protein Sci.* **1**: 1508-1521.

Blöchl, E., Rachel, R., Burggraf, S., Hafenbradl, D., Jannasch, H.W., and Stetter, K.O. (1997) *Pyrolobus fumarii*, gen. and sp. nov., represents a novel group of archaea, extending the upper temperature limit for life to 113°C. *Extremophiles* **1**: 14-21.

Blumentals, I.I., Robinson, A.S., and Kelly, R.M. (1990) Characterization of sodium dodecyl sulfate-resistant proteolyltic activity in the hyperthermophilic archaebacterium *Pyrococcus furiosus*. *Appl. Environm. Microbiol.* **56**: 1992-1998.

Bouyoub, A., Barbier, G., Forterre, P., and Labedan, B. (1996) The adenylosuccinate synthetase from the hyperthermophilic archaeon *Pyrococcus* species displays unusual structural features. *J. Mol. Biol.* **261**: 144-154.

Budgen, N., and Danson, M.J. (1986) Metabolism of glucose via a modified Entner-Douderoff pathway in the thermoacidophilic archaebacterium *Thermoplasma acidophilum*. *FEBS Lett.* **196**: 206-210.

Burlini, N., Magnani, P., Villa, A., Macchi, F., Tortora, P., and Guerritore, A. (1992) A heat-stable serine proteinase from the extreme thermophilic archaebacterium *Sulfolobus solfataricus*. *Biochim. Biophys. Acta* **1122**: 283-292.

Cannio, R., de Pascale, D., Rossi, M., and Bartolucci, S. (1994) Gene expression of a thermostable β-galactosidase in mammalian cells and its application in assays of eukaryotic promoter activity. *Biotechn. Appl. Biochem.* **19**: 233-244.

Carter, P., and Wells, J.A. (1988) Dissecting the catalytic triad of a serine protease. *Nature* **332**: 564-568.

Cavagnero, S., Zhou, Z.H., Adams, M.W.W., and Chan, S.I. (1995) Response of ruberdoxin from *Pyrococcus furiosus* to environmental changes: Implications for the origin of hyperthermostability. *Biochemistry* **34**: 9865-9873.

Chan, M.K., Mukund, S., Kletzin, A., Adams, M.W.W., and Rees, D.C. (1995) Structure of a hypethermophilic tungstopterin enzyme, aldehyde ferredoxin oxidoreductase. *Science* **267**: 1463-1469.

Cohen-Kupiec, R., Blank, C., and Leigh, J.A. (1997) Transcriptional regulation in Archaea: *in vivo* demonstration of a repressor binding site in a methanogen. *Proc. Natl. Acad. Sci. USA* **94**: 1316-1320.

Connaris, H., Cowan, D.A., and Sharp, R.J. (1991) Heterogeneity of proteinases from the hyperthermophilic archaebacterium *Pyrococcus furiosus*. *J. Gen. Microbiol.* **137**: 1193-1199.

Cowan, D.A., Smolenski, K.A., Daniel, R.M., and Morgan, H.W. (1987) An extremely thermostable extracellular proteinase from a strain of the archaebacterium *Desulfurococcus* growing at 88°C. *Biochem. J.* **247**: 121-133.

Daniel, R.M., Toogood, H.S., and Bergquist, P.L. (1995) Thermostable proteases. In: *Biotechnology and Genetic Engineering Reviews*. Vol. 13. pp 51-100.

Danson, M.J. (1988) Archaebacteria: the comparative enzymology of their central metabolic pathways. *Adv. Microbial Physiol.* **29**: 166-231.

Danson, M.J. (1989) Central metabolism of the archaebacteria: an overview. *Can. J. Microbiol.* **35**: 58-64.

De Rosa, M., Gambacorta, A., Nicoleus, B., Giardina, P., Poerio, E., and Buonocore, V. (1984) Glucose metabolism in the extreme thermoacidophilic archaebacterium *Sulfolobus solfataricus*. *Biochem. J.* **224**: 407-414.

Dedecker, B.S., O'Brien, R., Fleming, P.L., Geiger, J.H., Jackson, S.P., and Sigler, P.B. (1996) The crystal structure of a hyperthermophilic archaeal TATA-box binding protein. *J. Mol. Biol.* **264**: 1072-1084.

Eggen, H.I.L., Geerling, A., Watts, J., and de Vos, W.M. (1990) Characterization of pyrolysin, a hyperthermoactive serine protease from the archaebacterium *Pyrococcus furiosus*. *FEMS Microbiol. Lett.* 71: 17-20.

Eggen, R.I.L., Geerling, A.C.M., Voorhorst, W.G.B., Kort, R., and de Vos, W.M. (1994) Molecular and comparative analysis of the hyperthermostable *Pyrococcus furiosus* glutamate dehydrogenase and its gene. *Biocatalysis* **8**: 131-141.

Fiala, G., and Stetter, K.O. (1986) *Pyrococcus furiosus* sp. nov. represents a novel genus of marine heterotrophic archaebacteria growing optimally at 100°C. *Arch. Microbiol.* **145**: 56-61.

Fitz-Gibbon, S., Choi, A.J., Miller, J.H., Stetter, K.O., Simon, M.I., Swanson, R., and Kim, U. (1997) A fosmid-based genomic map and identification of 474 genes of the hyperthermophilic archaeon *Pyrobaculum aerophilum*. *Extremophiles* 1: 36-51.

Fontana, A. (1991) How nature engineers proteins thermostability. In *Life under extreme conditions: Biochemical adaptation*. Berlin, Springer. 89-113.

Fox, J.L (1997) Whole *E.coli*! Microbial sequences in log-phase growth. *ASM News*,**63**; 187-192.

Fersht, A.R., and Serrano, L. (1993) Principles of protein stability derived from protein engineering experiments. *Current Opinion in Structural Biology* **3**: 75-83.

Fujiwara, S., Okuyama, S., and Imanaka, T. (1996) The world of archaea: genome analysis, evolution and thermostable enzymes. *Gene* **179**: 165-170.

Fusek, M., Lin, X.L., and Tang, J. (1990) Enzymic properties of thermopsin. *J. Biol. Chem.* **265**: 1496-1501.

Gray, C.J. (1993) Stabilisation of enzymes with soluble additives.In: *Thermostability of enzymes*. Ed Gupta, M.N. Springer Verlag, NY, pp 124-145.

Grogan, D. (1991) Evidence that β-galactosidase of *Sulfolobus solfataricus* is only one of the several activities of a thermostable β-D-glycosidase. *Appl. Environ. Microbiol.* **57**: 1644-1649.

Gueguen, Y., Voorhorst, W.G.B., van der Oost, J., and de Vos, W.M. (1997) Molecular and biochemical characterization of an endo-β-1,3-glucanase of the hyperthermophilic archaeon *Pyrococcus furiosus*. *J. Biol. Chem. (in press)*

Gupta M.N. (1991) Thermostabilization of proteins. *Biotechn. Appl. Biochem.* **14**:1-11.

Halio, S.B., Blumentals, I., Short, S.A., Merrill, B.M., and Kelly, R. (1996) Sequence, expression in *Escherichia coli*, and analysis of the gene encoding a novel intracellular protease (PfpI) from the hyperthermophilic archaeon *Pyrococcus furiosus*. *J. Bacteriol* **178**: 2605-2612.

Hanzawa, S., Hoaki, T., Jannasch, H.W., and Maruyama, T. (1996) An extremely thermostable serine protease from a hyperthermophilic archaeum, *Desulfurococcus* strain SY, isolated from a deep-sea hydrothermal vent. *J. Mar. Biotechnol.* **4**: 121-126.

Harwood, V.J., Denson, J.D., Robinson-Bidle, K.A., and Schreier, H.J. (1997) Overexpression and characterization of a prolyl endopeptidase from the hyperthermophilic archaeon *Pyrocoocus furiosus*. *J. Bacteriol.* 179: 3613-3618.

Haseltine, C., Rolfsmeier, M., and Blum, P. (1996) The glucose effect and regulation of a a-amylase synthesis in the hyperthermophilic archaeon *Sulfolobus solfataricus*. *J. Bacteriol.* **178**: 945-950.

Heider, J., Mai, X., and Adams, M.W.W. (1996) Characterization of 2-ketoisovalerate ferredoxin oxidoreductase, a new and reversible coenzyme A-dependent enzyme involved in peptide fermentation by hyperthermophilic archaea. *J. Bacteriol.* **178**: 780-787.

Hennig, M., Darimont, B., Sterner, R., Kirschner, K., and Jansonius, J.N. (1995) 2.0 Å structure of indole-3-glycerol phosphate synthase from the hyperthermophile *Sulfolobus solfataricus*: possible determinants of protein stability. *Structure* 3: 1295-1306.

Henrissat, B. (1997) http://expasy.hcuge.ch/cgi-bin/lists?glycosid.txt

Henrissat B., and Bairoch A. (1996) Updating the sequence-based classification of glycosyl hydrolases. *Biochem. J.* **316**: 695-696.

Hensel, R., Laumann, S., Lang, J., Heumann, H., and Lottspeich, F. (1987) Characterization of two D-glyceraldehyde-3-phosphate dehydrogenases form the extremely thermophilic archaebacterium *Thermoproteus tenax*. *Eur. J. Biochem.* **170**: 325-333.

Hess, D., Krüger, K., Knappik, A., Palm, P., and Hensel, R. (1995) Dimeric 3-phosphoglycerate kinases from hyperthermophilic archaea: Cloning, sequencing and expression of the 3-phosphoglycerate kinase gene of *Pyrococcus furiosus* in *Escherichia coli* and characterization of the protein. Structural and functional comparison with the 3-phosphoglycerate kinase of *Methanothermus fervidus*. *Eur. J. Biochem.* **233**: 227-237.

Hethke, C., and Thomm, T. (1997) Unpublished data.

Horikoshi, K. (1997) A new microbial world-Extremophiles. *Extremophiles* 1: Editorial.

Kelly, R.M., and Adams, M.W.W. (1994) Metabolism in hyperthermophilic microorganisms. *Antonie Van Leeuwenhoek.* **66**: 247-270.

Ken, R., and Hackett, N.R. (1991) *Halobacterium* halobium strains lysogenic from phage ΦH contain a protein resembling coliphage repressors. J. *Bacteriol.* **173**: 955-960.

Kengen, S.W.M., de Bok, F.A.M., van Loo, N.D., Dijkema, C., Stams, A.J.M., and de Vos, W.M. (1994) Evidence for the operation of a novel Embden-Meyerhof pathway that involves ADP-dependent kinases during sugar fermentation by *Pyrococcus furiosus*. *J. Biol. Chem.* **269**: 17537-17541.

Kengen, S.W.M., Luesink, E.J., Stams, A.J.M., and Zehnder, A.J.B. (1993) Purification and characterization of an extremely stable β-glucosidase from the hyperthermophilic archaeon *Pyrococcus furiosus*. *Eur. J. Biochem.* **213**: 305-312.

Kengen, S.W.M., and Stams, A.J.M. (1994) An extremely thermostable β-glucosidase from the hyperthermophilic archaeon *Pyrococcus furiosus*. *Biocatlysis.* **11**: 79-88.

Kengen, S.W.M., Stams, A.J.M., and de Vos, W.M. (1996) Sugar metabolism of hyperthermophiles. *FEMS Microbiol. Rev.* **18**: 119-137.

Klingeberg, M., Friedrich, A., and Antranikian, G. (1992) Production of heat-stable proteases from thermophilic microorganisms and their application in the degradation of chicken feathers. *DECHEMA Biotechnological Conferences* 5: 173-176.

Klingeberg, M., Galunsky, B., Sjoholm, C., Kasche, V., and Antranikian, G. (1995) Purification and properties of highly thermostable, SDS resistant and stereospecific proteinase from the extreme thermophilic archaeon *Thermococcus stetteri*. *Appl. Environ. Microbiol.* **61**: 3098-3104.

Klingeberg, M., Hashwa, F., and Antranikian, G. (1991) Properties of extremely thermostable proteases from anaerobic hyperthermophilic bacteria. *Appl. Microbiol. Biotechnol.* **34**: 715-719.

Knapp, S., de Vos, W.M., Rice, D., and Ladenstein, R. (1997) Crystal structure of glutamate dehydrogenase from the hyperthermophilic eubacterium *Thermotoga maritima* at 3.0 Å resolution. *J. Mol. Biol.* **267**: 916-932.

Korndörfer, I., Steipe, B., Huber, R., Tomschy, A., and Jaenicke, R. (1995) The crystal structure of holo-glyceraldehyde-3-phosphate dehydrogenase from the hyperthermophilic bacterium *Thermotoga maritima* at 2.5 Å resolution. *J. Mol. Biol.* **246**: 511-521.

Kyrpides, and Ouzounis (1995) The eubacterial transcriptional activator Lrp is present in the archaeon *Pyrococcus furiosus*. *Trends Biochem. Sci.* **20**: 140-141.

Langer, D., Hain, J., Thuriaux, P., and Zillig, W. (1995) Transcription in *Archaea*: Similarity to that of *Eucarya*. *Proc. Natl. Acad. Sci. USA* **92**: 5768-5772.

Lebbink, J.H.G., Eggen, R.I.L., Geerling, A.C.M., Consalvi, V., Chiaraluce, R., Scandurra, R., and de Vos, W.M. (1995) Exchange of domains of glutamate dehydrogenase from the hyperthermophilic archaeon *Pyrococcus furiosus* and the mesophilic bacterium *Clostridium difficile*: effects on catalysis, thermoactivity and stability. *Prot. Engng.* **8**: 1287-1294.

Lebbink, J.H.G., Tuininga, J.E., van der Oost, J., and de Vos, W.M. (1997) Unpublished results.

Leuschner, C., and Antranikian, G. (1995) Heat-stable enzymes from extremely thermophilic and hyperthermophilic microorganisms. *World J. Microbiol. Biotechn.* **11**: 95-114.

Lin, X.L., and Tang, J. (1990) Purification characterization and gene cloning of thermopsine a thermostable acid protease from *Sulfolobus acidocaldarius*. *J. Biol. Chem.* **265**: 1490-1495.

Mai, X., and Adams, M.W.W. (1996) Purification and characterization of two reversible and ADP-dependent acetyl Coenzyme A synthetases from the hyperthermophilic archaeon *Pyrococcus furiosus*. *J. Bacteriol.* **178**: 5897-5903.

Mai, X.H., and Adams, M.W.W. (1994) Indolepyruvate ferredoxin oxidoreductase from the hyperthermophilic archaeon *Pyrococcus furiosus*: A new enzyme involved in peptide fermentation. *J. Biol. Chem.* **269**: 16726-16732.

Martins, L.O., and Santos, H. (1995) Accumulation of mannosylglycerate and di-*myo*-inositol-phosphate by *Pyrococcus furiosus* in response to salinity and temperature. *Appl. Environ. Microbiol.* **61**: 3299-3303.

Mayr, J., Lupas, A., Kellermann, J., Eckerskorn, C., Baumeister, W., and Peters, J. (1996) A hyperthermostable protease of the subtilisin family bound to the surface layer of the Archaeon *Staphylothermus marinus*. *Current Biology* 6: 739-749.

Moracci , M., Capalbo, L., Ciaramella, M., and Rossi. M. (1996) Identification of two glutamic acid residues essential for catalysis in the β-glycosidase from the thermoacidophilic archaeon *Sulfolobus solfataricus*. *Protein Engng.* **9**: 1191-1195.

Moracci, M., La Volpe, A., Pulitzer, J.F., Rossi, M., and Ciaramella, M. (1992) Expression of the thermostable β-galactosidase gene from the archaebacterium *Sulfolobus solfataricus* in *Saccharomyces cerevisiae* and characterization of a new inducible promoter for heterologous expression. *J. Bacteriol.* **174**: 873-882.

Muir, J.M., Russell, R.J.M., Hough, D.W., and Danson, M.J. (1995) Citrate synthase from the hyperthermophilic archaeon *Pyrococcus furiosus*. *Protein Engng.* **8**: 583-592.

Mukund, S., and Adams, M.W.W. (1995) Glyceraldehyde-3-phosphate ferredoxin oxidoreductase, a novel tungsten-containing enzyme with a potential glycolytic role in the hyperthermophilic archaeon *Pyrococcus furiosus*. *J. Biol. Chem.* **270**: 8389-8392.

Nucci, R., D'Auria, S., Febbraio, F., Vaccaro, C., Morana, A., De Rosa, M., and Rossi, M. (1995) A thermostable β-glycosidase from *Sulfolobus solfararicus*: temperature and SDS effects on its functional and structural properties. *Biotechn. Appl. Biochem.* **21**: 265-274.

Nucci, R., Moracci, M., Vaccaro, C., Vespa, N., and Rossi, M. (1993) Exoglucosidase activity and substrate specificity of the β-glycosidase isolated from the extreme thermophile *Sulfolobus solfataricus*. *Biotechn. Appl. Biochem.* **17**: 239-250.

Pedroni, P., Della Volpe, A., Galli, G., Mura, G.M., Pratesi, C., and Grandi, G. (1995) Characterization of the locus encoding the [Ni-Fe] sulfhydrogenase from the archaeon Pyrococcus furiosus: evidence for a relationship to bacterial sulfite reductases.*Microbiology* **141**:449-

Querol, E., Perex-Pons, J.A., and Mozo-Villarias, A. (1996) Analysis of protein conformation characterisitics related to thermostability. *Protein Engng.* 9: 265-271.

Rice, D.W., Engel,P.C., Ohshima, T., Robb, F.T., and Scandurra, R. (1996) Generic lessons on protein stability from studies on glutamate dehydrogenase. Thermophiles 96 Conference. p 54.

Robinson, K.A., Bartley, D.A., Robb, F.T., and Schreier, H.J. (1995) A gene from the hyperthermophile *Pyrococcus furiosus* whose deduced product is homologous to members of the prolyl oligopeptidase family of proteases. *Gene* **152**: 103-106.

Schäfer, T., and Schönheit, P. (1993) Gluconeogenesis from pyruvate in the hyperthermophilic archaeon *Pyrococcus furiosus* - involvement of reactions of the Embden-Meyerhof pathway. *Arch. Microbiol.* **159**: 354-363.

Schäfer, T., Selig, M., and Schönheit, P. (1993) Acetyl-CoA synthetase (ADP-forming) in *Archaea*, a novel enzyme involved in acetate formation and ATP synthesis. *Arch. Microbiol.* **159**: 72-83.

Schönheit, P., and Schäfer, T. (1995) Metabolism of hyperthermophiles. *World J. Microbiol. Biotechn.* **11**: 26-57.

Selig, M., Xavier, K., Santos, H., and Schönheit, P. (1997) Comparative analysis of Embden-Meyerhof and Entner-Douderoff glycolytic pathways in hyperthermophilic archaea and the bacterium *Thermotoga. Arch Microbiol.* **167**: 217-232.

Sensen, C.W., Klenk, H.-P., Singh, R.K., Allard, G., Chan, C.-Y., Lui, Q.Y., Penny, S.L., Young, F., Schenk, M.E., Gaasterland, T., Doolittle, W.F., Ragan, M.A., and Charlebois, R.L. (1996) Organizational characteristics and information content of an archaeal genome: 156 kb of sequence from *Sulfolobus solfataricus* P2. *Mol. Micriobiol.* **21**

Siebers, B., and Hensel, R. (1993) Glucose catabolism of the hyperthermophilic archaeum *Thermoproteus tenax*. *FEMS Microbiol. Lett.* **111**: 1-8.

Siezen, R.J., de Vos, W.M., Leunissen, J.A.M., and Dijkstra, B.W. (1991) Homology modelling and protein engineering strategy of subtilases, the family of subtilisin-like serine proteases. *Protein Engng.* **4**: 719-737.

Siezen, R.J., and Leunissen (1997) Subtilases: the superfamily of subtilisin-like serine proteases. *Protein Sci.* **6**: 501-523.

Stetter, K.O., Fiala, G., Huber, G., Huber, R., and Segerer, A. (1990) Hyperthermophilic microorganisms. *FEMS Microbiol. Rev.* **75**: 117-124.

Stetter, K.O. (1994) The lesson of Archaebacteria. In *Nobel symposium.* New York, Columbia University Press. 143-151.

Stolt, P., and Zillig, W. (1992) *In vivo* studies on the effects of immunity genes on early lytic transcription in the *Halobacterium salinarium* phage phi H. *Mol. Gen. Genet.* **235**: 197-204.

Sunna, A., Moracci, M., Rossi, M., and Antranikian, G. (1997) Glycosyl hydrolases from hyperthermophiles. *Extremophiles* **1**: 2-13.

Szilagyi, A., and Zavodsky, P. (1995) Structural basis for the extreme thermostability of D-glyceraldehyde-3-phosphate dehydrogenase from *Thermotoga maritima*: analysis based on homology modelling. *Protein Engng.* **8**: 779-789.

Tomschy, A., Böhm, G., and Jaenicke, R. (1994) The effect of ion pairs on the thermal stability of D-glyceraldehyde-3-phosphate dehydrogenase from the hyperthermophilic bacterium *Thermotoga maritima. Prot. Engng* **7**: 1471-1478.

Tutino, M.L., Scarano, G., Marino, G., Sannia, G., and Cubellis, M.V. (1993) Tryptophan biosynthesis genes *trpEGC* in the thermoacidophilic archaebacterium *Sulfolobus solfataricus*. *J. Bacteriol.* **175**: 299-

Teng, Q., Zhou, Z.H., Smith, W.T., Busse, S.C., Howard, J.B., Adams, M.W.W., and La Mar, G.N. (1994) Solution of ^{1}H NMR determination of secondary structure for the three-iron form of ferredoxin from the hyperthermophilic archaeaon *Pyrococcus furiosus*. *Biochemistry* **33**: 6316-6326.

Thomm, M. (1996) Archaeal transcription factors and their role in transcription initiation. *FEMS Microbiol. Rev.* **18**: 159-171.

Van der Oost, J., Ciaramella, M., Moracci, M., Pisani, F.M., Rossi, M., de Vos, W.M. (1997) Molecular biology of hyperthermophilic Archaea. *Adv. Biochem Engng./Biotechnol.* (in press)

Van der Oost, J., Schut, G., Kengen, S.W.M., Hagen, W.R., Thomm, M., De Vos, W.M. (1997) The ferredoxin-dependent conversion of glyceraldehyde-3-phosphate in the hyperthermophilic archaeon Pyrococcus furiosus represents a novel site of glycolytic regulation. (in preparation).

Verhees, C., van der Oost, J., and de Vos, W.M. (1997) (Unpublished results).

Vieille, C., and Zeikus, J.G. (1996) Thermozymes: identifying molecular determinants of protein structural and functional stability. *Trends Biotechn.* **14**: 183-190.

Völkl, P., Markiewicz, P., Stetter, K.O., and Miller, J.H. (1995) The sequence of a subtilisin-type protease (aerolysin) from the hyperthermophilic archaeum *Pyrobaculum aerophilum* reveals sites important to thermostability. *Protein Science* **3**: 1329-1340.

Voorhorst, W.G.B., Eggen, R.I.L., Geerling, A.C.M., Platteeuw, C., Siezen, R.J., and de Vos, W.M. (1996) Isolation and characterization of the hyperthermostable serine protease, pyrolysin, and its gene from the hyperthermophilic archaeon *Pyrococcus furiosus*. *J. Biol. Chem.* **271**: 20426-20431.

Voorhorst, W.G.B., Eggen, R.I.L., Luesink, E.J., and de Vos, W.M. (1995) Characterization of the *celB* gene coding for β-glucosidase from the hyperthermophilic archaeon *Pyrococcus furiosus* and its expression and site-directed mutation in *Escherichia coli*. *J. Bacteriol.* 177: 7105-7111.

Voorhorst, W.G.B., Gueguen, Y., Schut, G., Dahlke, I., Thomm, M., van der Oost, J., and de Vos, W.M. (1997) Transcriptional regulation in the hyperthermophilic archaeon *Pyrococcus furiosus*: Coordinated expression of divergently transcribed genes in response to β-linked glucose polymers. *Submitted for publication.*

Voorhorst, W.G.B., Warner, A., de Vos, W.M., and Siezen, R.J. (1997) Homology modelling of two subtilisin-like serine proteases form the hyperthermophilic archaea *Pyrococcus furiosus* and *Thermococcus stetteri*. *Protein Engng.* **10**:

Wettach, J., Gohl, H.P., Tschochner, H., and Thomm, M. (1995) Functional interaction of yeast and human TATA-binding proteins with an archaeal RNA polymerase and promoter. *Proc. Natl. Acad. Sci. USA* **92**: 472-476.

Woese, C.R., Kandler, O., and Wheelis, M. (1990) Towards a natural system of organisms: proposal for the domains *Archaea, Bacteria* and *Eucarya*. *Proc. Natl. Acad. Sci. USA* **87**: 4576-4579.

Yip, K.S.P., Stillman, T.J., Britton, K.L., Artymiuk, P.J., Baker, P.J., Sedelnikova, S.E., Engel, P.C., Pasquo, A., Chiaraluce, R., Consalvi, V. *et al.* (1995) The structure of *Pyrococcus furiosus* glutamate dehydrogenase reveals a key role for ion-pair networks in maintaining enzyme stability at extreme temperatures. *Structure* **3**: 1147-1158.

Zwickl, P., Fabry, S., Bogedain, C., Haas, A., and Hensel, R. (1990) Glyceraldehyde-3-phosphate dehydrogenase from the hyperthermophilic archaebacterium *Pyrococcus woesei*: characterization of the enzyme, cloning and sequencing of the gene, and expression. *J. Bacteriol.* **172**: 4329-4338.

Zillig, W., Palm, P., Klenk, H-P., Langer, D., Hüdepohl, U., Hain, J., Lanzendörfer, M., and Holz, I. (1993) Transcription in Archaea. In *The biochemistry of Archaea (Archaebacteria)* M. Kates, Kushner, D.J., and Matheson, A.T., Elsevier, NY, p367-391.

Chapter 2

Characterization of the *celB* Gene Coding for β-Glucosidase from the Hyperthermophilic Archaeon *Pyrococcus furiosus* and Its Expression and Site-Directed Mutation in *Escherichia coli*

Wilfried G.B. Voorhorst, Rik I.L. Eggen, Evert J. Leusink, and Willem M. de Vos

reprinted with permission from the **Journal of Bacteriology**, 1995, 7105-7111

JOURNAL OF BACTERIOLOGY, Dec. 1995, p. 7105–7111 Vol. 177, No. 24
0021-9193/95/$04.00+0

Characterization of the *celB* Gene Coding for β-Glucosidase from the Hyperthermophilic Archaeon *Pyrococcus furiosus* and Its Expression and Site-Directed Mutation in *Escherichia coli*

WILFRIED G. B. VOORHORST, RIK I. L. EGGEN,† EVERT J. LUESINK,‡ AND WILLEM M. DE VOS*

Bacterial Genetics Group, Department of Microbiology, Wageningen Agricultural University, 6703 CT Wageningen, The Netherlands

Received 14 July 1995/Accepted 12 October 1995

The *celB* gene encoding the cellobiose-hydrolyzing enzyme β-glucosidase from the hyperthermophilic archaeon *Pyrococcus furiosus* has been identified, cloned, and sequenced. The transcription and translation initiation sites of the *celB* gene have been determined, and archaeal control sequences were identified. The *celB* gene was overexpressed in *Escherichia coli*, resulting in high-level (up to 20% of total protein) production of β-glucosidase that could be purified by a two-step purification procedure. The β-glucosidase produced by *E. coli* had kinetic and stability properties similar to those of the β-glucosidase purified from *P. furiosus*. The deduced amino acid sequence of CelB showed high similarity with those of β-glycosidases that belong to glycosyl hydrolase family 1, implicating a conserved structure. Replacement of the conserved glutamate 372 in the *P. furiosus* β-glucosidase by an aspartate or a glutamine led to a high reduction in specific activity (200- or 1,000-fold, respectively), indicating that this residue is the active site nucleophile involved in catalysis above 100°C.

The most extensively studied representative of the hyperthermophilic organisms that have an optimum growth temperature above 85°C is *Pyrococcus furiosus* (12). *P. furiosus* is able to grow on a wide range of substrates, including complex polymers such as starch, glycogen, peptone, and casein or simple carbon compounds like cellobiose, maltose, and pyruvate (12, 20, 33). The main fermentation products are CO_2 and H_2 or alanine, the latter acting as an alternative electron sink (22). The disaccharides cellobiose and maltose are hydrolyzed by intracellular glucosidases (5, 20). The generated glucose was proposed to be further metabolized via a nonphosphorylated Entner-Doudoroff pathway (27, 33). However, recently it was discovered that sugars are fermented by *P. furiosus* via an Embden-Meyerhof pathway that involves two ADP-dependent kinases (19).

Characterization of proteins from hyperthermophiles revealed that they are extremely thermostable and may have an optimum temperature of catalysis that exceeds the maximum growth temperature of their host (1, 18). In addition to their remarkable thermostability, proteins from hyperthermophiles are often found to be highly resistant to chemical denaturation and to degradation by proteases (13). One of the most thermostable enzymes identified up to now is the β-glucosidase from *P. furiosus*, with a half-life of 85 h at 100°C (20). During growth on cellobiose, β-glucosidase can make up to 5% of the total cell protein of *P. furiosus* and is involved in the hydrolysis of the β-1,4-glycosidic bond between the two glucose moieties of the disaccharide (20). In addition, β-glucosidases constitute a group of well-studied enzymes among members of all three domains of life, *Eucarya*, *Bacteria*, and *Archaea* (16, 17, 35). Therefore, the pyrococcal β-glucosidase is a suitable model enzyme for the molecular characterization of structure-function relations of hyperthermostable enzymes. To study these relations by protein engineering it is required to express the gene and produce a functional enzyme in an accessible heterologous host, since for hyperthermophilic members of the *Archaea* no genetic systems are presently available.

Here we report the isolation, cloning, and sequencing of the β-glucosidase (*celB*) gene from *P. furiosus*. Control regions involved in *celB* transcription initiation were identified, and the *celB* gene was functionally overexpressed in *Escherichia coli*. Subsequently, the CelB was purified and compared to the original enzyme and the active site nucleophile was identified by protein engineering.

MATERIALS AND METHODS

Organisms, media, and plasmids. *P. furiosus* (DSM 3638) was cultured at 98°C in synthetic seawater as previously described (20). *E. coli* TG1 (15), MC1061 (3), and JM109(DE3) (34), obtained from Pharmacia (Uppsala, Sweden), were grown in L broth and handled as described previously (31). The phagemids pTZ18R and pTZ19R and the expression vector pTTQ19 were obtained from Pharmacia.

Purification of β-glucosidase from *P. furiosus*. β-Glucosidase was purified from *P. furiosus* cells grown on cellobiose as previously described (20), and its NH_2-terminal amino acid sequence was determined by Edman degradation using an Applied Biosystems model 477A (gas-phase amino acid Sequenator) (Applied Biosystems, Foster City, Calif.) (courtesy of SON protein sequence facility, Leiden, The Netherlands).

Cloning of the β-glucosidase gene. On the basis of the NH_2-terminal amino acid sequence of the purified β-glucosidase, two oppositely orientated PCR oligonucleotides were designed: 5′-AARTTYCCNAARAAYTTYATGTTYGG (OL92-9), based on amino acid residues 2 to 9, and 5′-CCCATYTCRAAYT GRAA (OL92-10), based on amino acid residues 16 to 21 (where N is A, C, G, or T; R is A or G; and Y is C or T). Subsequently, part of the *celB* gene was amplified by PCR on a DNA Thermal Cycler (Perkin Elmer Cetus, Norwalk, Conn.), using 175 ng of chromosomal DNA of *P. furiosus* isolated as described previously (10) and 100 ng of both OL92-9 and OL92-10 in a final volume of 100 μl. Thirty cycles (1 min at 94°C, 2 min at 45°C, and 3 min at 72°C) were followed by a final incubation at 72°C for 7 min. To visualize the expected small product (59 bp) on a sequencing gel, [α-^{32}P]dATP (110 TBq/mmol) was added during amplification. A product with the expected size was used as probe in a Southern hybridization with chromosomal DNA from *P. furiosus*. Hybridizing fragments were isolated from agarose gels using Gene Clean (BIO 101, La Jolla, Calif.) and cloned in *E. coli* by established methods (31).

Nucleotide sequence analysis. Phagemids pTZ18R and pTZ19R were used to

* Corresponding author. Mailing address: Department of Microbiology, Wageningen Agricultural University, Hesselink van Suchtelenweg 4, 6703 CT Wageningen, The Netherlands. Phone: 31 3174 83100. Fax: 31 3174 83829. Electronic mail address: WILLEM.DEVOS@ALGEMEEN.MICR.WAU.NL.

† Present address: EAWAG, Dübendorf, Switzerland.

‡ Present address: NIZO, Ede, The Netherlands.

produce single-stranded DNA for sequence analysis as described by the supplier (Pharmacia). To obtain overlapping sequences, deletions were generated with exonuclease III (Life Technologies B.V., Breda, The Netherlands) (31). Sequencing was carried out on single-stranded DNA template using the dideoxy terminator method (32).

Computer analysis of nucleotide and protein sequences and alignments of these sequences were done with the PC/GENE program version 5.01 (IntelliGenetics Inc., Mountain View, Calif.) and the GCG package version 7.0 (7) at the CAOS/CAMM Centre of the University of Nijmegen.

RNA isolation and primer extension. Total RNA of *P. furiosus* was isolated from cells grown on cellobiose using guanidinium isothiocyanate and β-mercaptoethanol (4). For this purpose, cells were resuspended in denaturation buffer (4 M guanidinium isothiocyanate, 42 mM sodium citrate, 0.83% *N*-lauryl sarcosine, 0.2 mM β-mercaptoethanol) followed by the addition of 0.1 volume of 2 M ammonium acetate (pH 4) and 1 volume of phenol-chloroform-isoamyl alcohol (25:24:1). The mixture was incubated on ice for 15 min, and phase separation was obtained by centrifugation for 20 min (10,000 × g, 4°C). The aqueous phase was removed and supplemented with an equal volume of isopropanol, and RNA was precipitated at −20°C. The pelleted RNA was resuspended in denaturation buffer, precipitated again, washed with ice-cold 75% ethanol, dried, and resuspended in water. This purified RNA was used as a template in primer extension experiments with oligonucleotide BG72 (5′-CCAAGAATATCCAAACATGA AG), as previously described (10).

Mutagenesis of the *celB* gene. Mutations were introduced in the *celB* gene using the CLONE AMP site-directed mutagenesis system (Life Technologies) (28). The mutagenic oligonucleotides (BG 95 [5′-CCAAUGAUAAUUACAN AVAACGGUAUG] and BG 79 [5′-GGCCAUACCGUUBUNUGUAA], where N is A, C, G, or U; V is G, C, or A; and B is G, C, or U) were each used in combination with a dU-*lacZ* primer during a PCR amplification with 5 ng of *Sac*I-linearized pLUW503 for 30 cycles in a total volume of 100 μl. The PCR products were combined, and the pAMP cloning vector was added. After removal of the uracil residues by uracil DNA glycosylase, the DNA was introduced by transformation into the supplied *E. coli* DH5α and the resulting recombinant plasmids were analyzed by sequencing the region containing the mutation. From the plasmids containing a mutated *celB* gene, the 190-bp *Apa*I-*Aff*II fragment coding for residues 357 to 418 was cloned into *Apa*I-*Aff*II-digested pLUW503 and transformed into *E. coli* JM109(DE3).

Characterization of β-glucosidase produced in *E. coli*. *E. coli* TG1 or *E. coli* JM109(DE3) harboring either the wild-type or mutated *celB* gene was grown overnight in 5 ml of L broth with 100 μg of ampicillin per ml. If appropriate, 3 mM isopropyl-β-D-thiogalactopyranoside was added. Cells were harvested by centrifugation, concentrated 15× in citrate buffer (50 mM sodium citrate, pH 6.5), and disrupted by sonication. The cell debris was pelleted by centrifugation, and the resulting supernatant was used as cell extract. The β-glucosidase activity was determined using β-D-glucopyranoside-*p*-nitrophenyl (Boehringer GmbH, Mannheim, Germany) as a substrate (20). The kinetic parameters, temperature optimum, and thermostability of the purified β-glucosidase were determined by using the same synthetic substrate and done as previously described (20). Routine measurements were performed on a Beckman DU 7500 spectrophotometer with supplied temperature controller at 90°C (Beckman Instruments, Inc. Fullerton, Calif.). For the thermostability assay, β-glucosidase was incubated at 100°C at a concentration of 0.5 mg/ml in sealed glass vials.

Electrophoretic analysis of protein samples was done by sodium dodecyl sulfate-polyacrylamide gel electrophoresis (SDS-PAGE) by the method of Laemmli (24). Protein samples for SDS-PAGE were prepared by heating for 10 min at 110°C in an equal volume of sample buffer (0.1 M Tris-HCl, 5% SDS, 0.9% 2-mercaptoethanol, 20% glycerol, pH 6.8).

Overproduction of pyrococcal β-glucosidase in *E. coli* and its purification. *E. coli* MC1061 harboring pLUW510 was grown in L broth with 300 μg of ampicillin per ml to an optical density at 600 nm of 0.5. The *tac* promoter was subsequently induced by the addition of 3 mM isopropyl-β-D-thiogalactopyranoside, and the culture was incubated for another 10 h. Cells were harvested by centrifugation, resuspended in P buffer (50 mM sodium phosphate, pH 7.5), and broken by using a French press. RNase (10 μg/ml), DNase (10 μg/ml), and $MgCl_2$ (10 mM) were added, and after a 30-min incubation at 37°C the mixture was transferred to 80°C and incubated for another 45 min. During the latter incubation the majority of the *E. coli* proteins were denatured and could be removed by centrifugation. The supernatant was subsequently dialyzed against P buffer, loaded onto a Q-Sepharose Fast Flow (Pharmacia) column (5.0 by 2.5 cm) preequilibrated with the same buffer and eluted by a linear gradient of NaCl (0 to 1 M in P buffer). Activity fractions that eluted around 0.6 M NaCl were pooled and dialyzed against P buffer. The NH_2-terminal amino acid sequence of the CelB produced in *E. coli* was determined by Edman degradation using a gas-phase amino acid Sequenator.

Nucleotide sequence accession number. The nucleotide sequence reported has been submitted to the GenBank/EMBL Data Bank with the accession number U37557.

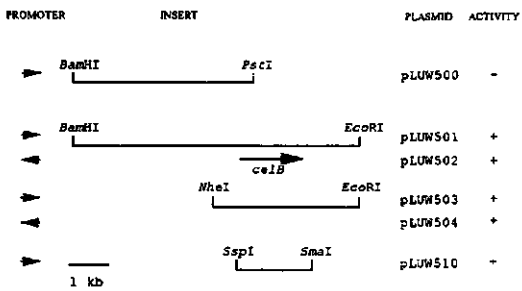


FIG. 1. Physical map of plasmids containing *celB*. The orientation of the vector-located *lacUV* promoter is shown. In addition, thermoactive β-glucosidase activity in cell extracts of *E. coli* harboring the respective plasmids is indicated: +, activity of approximately 4 to 20 U/mg protein; −, no detectable activity.

RESULTS

Cloning and expression of the *celB* gene encoding β-glucosidase from *P. furiosus*. The *P. furiosus* β-glucosidase was purified to homogeneity from cells grown on cellobiose, and its NH_2-terminal 23-amino-acid sequence showed the sequence NH_2-[M/A]KFPKNFMFGYXXSGFQFEMGLP (X is nonidentified residue). Based on amino acid residues 2 to 9 and 16 to 21, two degenerate oligonucleotides (OL92-9 and OL92-10) were designed and used as primers in a PCR reaction with *P. furiosus* DNA. An amplified fragment with the expected size of 59 bp was obtained and subsequently used as a probe in Southern hybridization with chromosomal DNA from *P. furiosus* digested with various restriction endonucleases. A *Bam*HI-*Pst*I fragment of 4.8 kb hybridized strongly and was cloned into pTZ19R digested with *Bam*HI and *Pst*I, resulting in pLUW500 (Fig. 1). Sequence analysis showed that this fragment contained an incomplete open reading frame comprising the 5′ end of the β-glucosidase gene, designated *celB* (see below).

To obtain a fragment containing the complete *celB* gene, the 4.8-kb *Bam*HI-*Pst*I fragment was used as a probe in a Southern hybridization with DNA from *P. furiosus*. A hybridizing 7.7-kb *Bam*HI-*Eco*RI fragment was isolated and cloned into a *Bam*HI-*Eco*RI-digested pTZ19R resulting in pLUW501 (Fig. 1). A cell extract of *E. coli* TG1 harboring pLUW501 showed β-glucosidase activity at 90°C, indicating that the *P. furiosus* gene for β-glucosidase had been cloned and was functionally expressed (Fig. 1). The 7.7-kb insert of pLUW501 could be reduced in size by deleting a 4.2-kb *Xba*I (multiple cloning site)-*Nhe*I (insert) fragment, resulting in pLUW503, that in *E. coli* TG1 was able to produce a functional β-glucosidase (Fig. 1).

Cloning of the inserts of pLUW501 and pLUW503 in pTZ18R resulted in plasmids pLUW502 and pLUW504, respectively, that in *E. coli* TG1 showed similar β-glucosidase activities (4 to 7 U/mg of protein; Fig. 1). This indicates that expression of the *celB* gene is not controlled by the vector-located *lac* promoter in these plasmids and that the region upstream of the *celB* gene is involved in transcription initiation in *E. coli*.

Nucleotide sequence analysis of the *celB* gene. The nucleotide sequence of the insert of pLUW503 was determined in both directions and showed the presence of an open reading frame of which the first ATG codon is preceded by a consensus, purine-rich ribosome binding site (Fig. 2). The sequence starting from this ATG initiation codon to the TAG stop codon includes 1,416 bp and could encode a 472-amino-acid protein

```
                                                    SspI                                    ↓ ******
-121 TTTTGAAGAGCCTTGAGCCTCCTCAAAAGCAGTTAACATATATCTATCTTCCCACATTTATAATTGCATGGGAAATATTATAAATCACAATATCAAAATATAAAGCTAGAGGTGGAAAGT

   1 ATGAAGTTCCCAAAAAACTTCATGTTTGGATATTCTTGGTCTGGTTTCCAGTTTGAGATGGGACTGCCAGGAAGTGAAGTGGAAAGCGACTGGTGGGTGTGGGTTCACGACAAGGAGAAC
   1 M K F P K N F M F G Y S W S G F Q F E M G L P G S E V E S D W W V W V H D K E N

 121 ATAGCATCAGGTCTAGTAAGTGGAGATCTACCAGAGAACGGCCCAGCATATTGGCACCTCTATAAGCAAGATCATGACATTGCAGAAAAGCTAGGAATGGATTGTATTAGAGGTGGCATT
  41 I A S G L V S G D L P E N G P A Y W H L Y K Q D N D I A E K L G M D C I R G G I

 241 GAGTGGGCAAGAATTTTTCCAAAGCCAACATTTGACGTTAAAGTTGATGTGGAAAAGGATGAAGAAGGCAACATAATTTCCGTAGACGTTCCAGAGAGTACAATAAAAGAGCTAGAGAAA
  81 E W A R I F P K P T F D V K V D V E K D E E G N I I S V D V P E S T I K E L E K

 361 ATTGCCAACATGGAGGCCCTTGAACATTATGCGAAGATTTACTCAGACTGGAAGGAGAGGGGCAAAACCTTCATATTAAACCTCTACCACTGGCCTCTTCCATTATGGATTCATGACCCA
 121 I A N M E A L E H Y A R I Y S D W K E R G K T F I L N L Y H W P L P L W I H D P

                  PstI
 481 ATTGCAGTAAGGAAACTTGGCCCGGATAGGGCTCCTGCAGGATGGTTAGATGAGAAGACAGTGGTAGAGTTTGTGAAGTTTGCCGCCTTCGTTGCTTATCACCTTGATGACCTCGTTGAC
 161 I A V R K L G P D R A P A G W L D E K T V V E F V K F A A F V A Y H L D D L V D

 601 ATGTGGAGCACAATGAACGAACCAAACGTAGTCTACAATCAAGGTTACATTAATCTACGTTCAGGATTTCCACCAGGATATCTAAGCTTTGAAGCAGCAGAAAAGGCAAAATTCAACTTA
 201 M W S T N N E P N V V Y N Q G Y I N L R S G F P P G Y L S F E A A E K A K F N L

 721 ATTCAGGCTCACATCGGAGCATATGATGCCATAAAAGAGTATTCAGAAAAATCCGTGGGAGTGATATACGCCTTTGCTTGGCACGATCCTCTAGCGGAGGAGTATAAGGATGAAGTAGAG
 241 I Q A H I G A Y D A I K E Y S E K S V G V I Y A F A W H D P L A E E Y K D E V E

 841 GAAATCAGAAAGAAAGACTATGAGTTTGTAACAATTCTACACTCAAAAGGAAAGCTAGACTGGATCGGCGTAAACTACTACTCCAGGCTGGTATATGGAGCCAAAGATGGACACCTAGTT
 281 E I R K K D Y E F V T I L H S K G K L D W I G V N Y Y S R L V Y G A K D G H L V

                                                                                  AflII
 961 CCTTTACCTGGATATGGATTTATGAGTGAGAGAGGAGGATTTGCAAAGTCAGGAAGACCTGCTAGTGACTTTGGATGGGAAATGTACCCAGAGGGCCTTGAGAACCTTCTTAAGTATTTA
 321 P L P G Y G F M S E R G G F A K S G R P A S D F G W E M Y P E G L E N L L K Y L

1081 AACAATGCCTACGAGCTACCAATGATAATTACAGAGAACGGTATGGCCGATGCAGCAGATAGATACAGGCCACACTATCTCGTAAGCCATCTAAAGGCAGTTTACAATGCTATGAAAGAA
 361 N N A Y E L P M I I T E N G M A D A A D R Y R P H Y L V S H L K A V Y N A M K E

                                  *
                                 ApaI
1201 GGTGCTGATGTTAGAGGGTATCTCCACTGGTCTCTAACAGACAACTACGAATGGGCCCAAGGGTTCAGGATGAGATTTGGATTGGTTTACGTGGATTTCGAGACAAAGAAGAGATATTTA
 401 G A D V R G Y L H W S L T D N Y E W A Q G F R M R F G L V Y V D F E T K K R Y L

1321 AGGCCAAGCGCCCTGGTATTCAGAGAAATAGCCACTCAAAAAGAAATTCCAGAAGAATTAGCTCACCTCGCAGACCTCAAATTTGTTACAAGAAAGTAGCCATTTCATTTTTTTCTTTGTT
 441 R P S A L V F R E I A T Q K E I P E E L A H L A D L K F V T R K *

1441 TTTTAACTCATATTTTTGGGAAAATAAGATCATAAAAGCTTAACAAGAGAAACCATGCCATAAAAATAGCAAAGATTGATAGCCCTCAATTTCCATGGATTCTACATACATATGGATGAT
     ---------------->     <----------------

1561 CCGTCATAAAGTTTATAAATCCATAGCAAGGTGATTTTATTGGTAGTATCCC
```

FIG. 2. Nucleotide sequence of the *P. furiosus celB* gene and its flanking regions. The deduced amino acid sequence of the open reading frame corresponding to CelB is depicted and relevant restriction sites are indicated. The transcription initiation site is indicated by the arrow. Regulatory sequences are indicated as follows: the box A of *celB* and the following open reading frame (encoding $tRNA^{Phe}$) (37) are underlined, the ribosome binding site is indicated by asterisks, and the NH_2-terminal amino acid residues of the purified β-glucosidase are underlined. In addition, T stretches downstream of the *celB* gene are overlined, and the inverted repeat is shown by the oppositely oriented arrows.

that is homologous to various β-glucosidases (see below). Since the deduced amino acid sequence of residues 1 to 23 is identical to that of the NH_2-terminal end of the purified β-glucosidase (Fig. 2), the coding region represents the *P. furiosus celB* gene. The calculated molecular mass of the deduced primary sequence is 54,580 Da, which is in agreement with the apparent molecular mass of 58 kDa determined by SDS-PAGE for the β-glucosidase purified from *P. furiosus* (20).

Transcriptional organization of the *celB* gene. Sequence analysis of the upstream region of the *celB* gene revealed a putative box A, the archaeal equivalent of the TATA box that directs transcription (Fig. 2) (38). To localize the *celB* transcription initiation site, primer extension was performed on total RNA isolated from *P. furiosus* cells grown on cellobiose. This resulted in an extension product with a size that corresponded with transcription initiation at an adenine residue located at position −14 that was located 25 nucleotides downstream of box A (Fig. 2 and 3). Termination of transcription may occur downstream of the *celB* gene, where three T-rich stretches are present, including the termination consensus sequence of at least four successive T residues (Fig. 2) (6). The last two of these T stretches are included in the left part of an inverted repeat.

Primary sequence comparison. Comparison of the deduced primary sequence of *celB* with β-glycosidases present in the SWISS PROT database (release 30) showed the highest similarities (69 to 44%) with representatives of glycosyl hydrolase family 1 and to a lower extent with members of family 3 (30% or lower) as defined by Henrissat (16). The CelB sequence was aligned with representatives of the family 1 of glycosyl hydrolases, which includes glucosidases, galactosidases, and 6-phospho-glycosidases of different origins (Fig. 4). Among the large number of conserved residues are the glutamate 372 and glu-

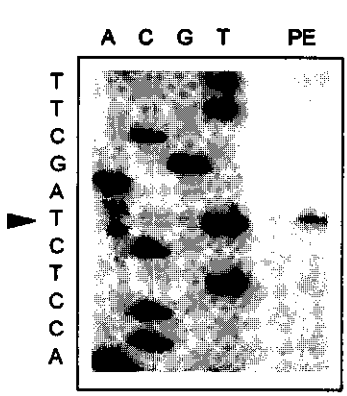


FIG. 3. Identification of the transcription initiation site of the *celB* gene by primer extension. Primer-extended product (PE) was electrophoresed in parallel with the sequence ladder generated with the same primer on the noncoding strand of *celB*.

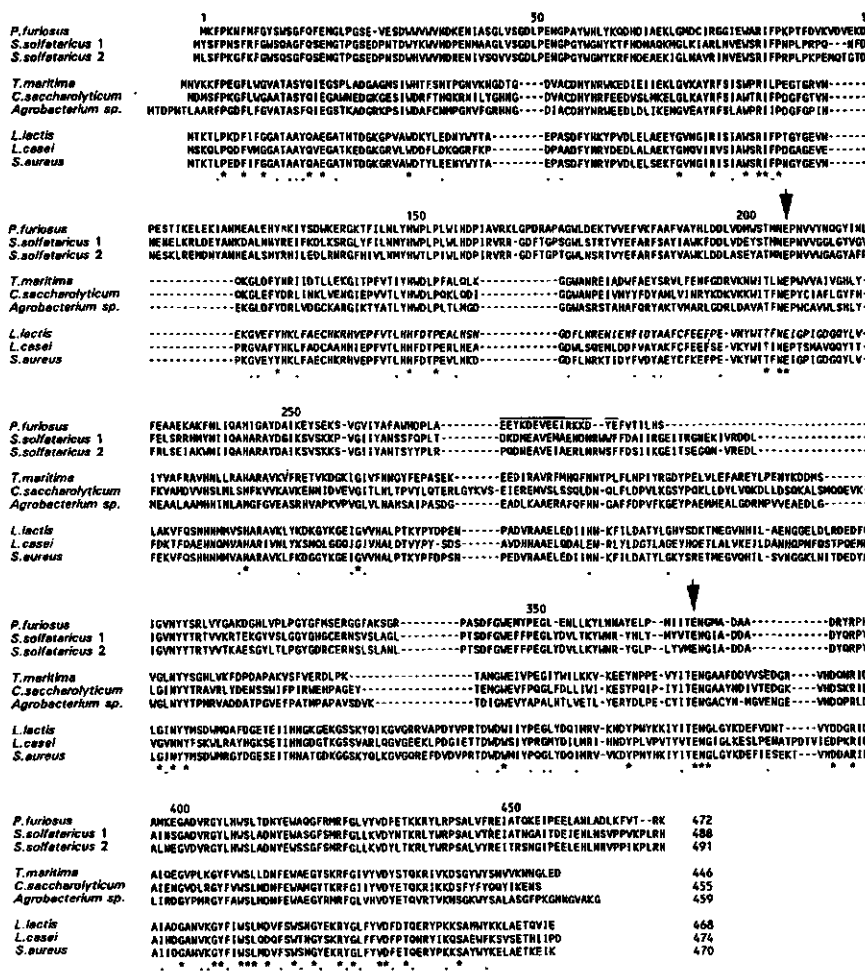


FIG. 4. Alignment of *P. furiosus* CelB with other members of family 1 of glycosyl hydrolases. Sequences were deduced from the following accession numbers: *Sulfolobus solfataricus* 1 β-galactosidase (P22498); *S. solfataricus* 2 β-galactosidase (P14288); *Thermotoga maritima* β-glucosidase; *Caldocellum saccharolyticum* β-glucosidase (SO3813); *Agrobacterium* sp. β-glucosidase (A28673); *Lactococcus lactis* 6-phospho-β-galactosidase (B37168); *Lactobacillus casei* 6-phosphogalactosidase (A29897); and *Staphylococcus aureus* 6-phospho-β-galactosidase (A27233). The numbering of the *P. furiosus* CelB is indicated. Conserved residues are indicated by an asterisk below the alignment, while dots represent amino acids with the same characteristics. The arrows above the sequence indicate the conserved glutamate residues that have been identified as acid-base catalyst and active site nucleophile (39, 40). The highly charged region (residue 273 to 288) found within the CelB sequence is overlined.

tamate 207 in the CelB sequence (Fig. 4). Glutamate 372 corresponds with the catalytically active nucleophile of the glycosyl hydrolases from *Agrobacterium* sp. and *Staphylococcus aureus*, while glutamate 207 has been identified as the acid-base catalyst in the staphylococcal enzyme (39, 40).

Purification and characterization of the *P. furiosus* β-glucosidase overproduced in *E. coli*. To allow for a biochemical characterization of the *P. furiosus* CelB produced in *E. coli*, it was of interest to purify it to homogeneity from a highly overproducing strain. In the case of the pTZ derivatives pLUW501, pLUW502, pLUW503, and pLUW504, the level of *celB* expression in *E. coli* was relatively low (4 to 20 U/mg of β-glucosidase activity) and not controlled by the *lac* promoter of the vector (Fig. 1). To improve and control the expression level of *celB* in *E. coli*, a 1.9-kb *Ssp*I-*Sma*I fragment of pLUW503 was inserted into the *Sma*I site of the expression vector pTTQ19, resulting in pLUW510 with the *celB* gene under control of the vector-located *tac* promoter. Upon induction with isopropyl-β-D-thiogalactopyranoside the level of CelB production in *E. coli* MC1061 harboring the pLUW510 could be increased up to 200 U/mg of total protein. Using this expression system, CelB could be produced up to 20% of the total cell protein. All β-glucosidase was soluble and no indications for the formation of inclusion bodies were found. The heterologous CelB could

easily be purified to homogeneity by a two-step purification procedure that involved a thermoincubation for 45 min at 80°C inducing denaturation of the majority of the host proteins, followed by an anion-exchange chromatography step (data not shown).

The kinetic and stability properties of the CelB produced by *E. coli* were compared with those from CelB purified from *P. furiosus*. This comparison showed that after a 30-h incubation at 100°C both enzymes retained 80 to 90% of the initial activity. In addition, the enzyme from both sources had the same temperature optimum (105°C) and showed identical K_m using β-D-glucopyranoside-*p*-nitrophenyl as substrate (0.39 and 0.41 mM for *P. furiosus* and *E. coli*, respectively) and a somewhat higher V_{max}, which may be explained by the rapid purification method, for the *E. coli*-produced enzyme (1,169 and 828 μmol/min/mg of protein for *E. coli* and *P. furiosus*, respectively). Moreover, the CelB produced in *E. coli* showed a linear Arrhenius plot irrespective of whether the extract had been heated (data not shown). An activation energy of 45 kJ/mol could be calculated for the *E. coli*-purified CelB similar to that reported previously for the pyrococcal CelB (20). Finally, analysis of the N-terminal residues of the CelB produced in *E. coli* showed that these were identical to that of the pyrococcal enzyme (data not shown). These results imply that CelB produced in *E. coli* is correctly folded and cannot be distinguished from that from *P. furiosus*, allowing protein engineering of this hyperthermostable enzyme.

Site-directed mutation of the *P. furiosus* β-glucosidase. To investigate whether the conserved glutamate at position 372 in CelB is the catalytically active nucleophile, it was replaced by an aspartate or a glutamine residue analogous to the approach used for glycosyl hydrolases of mesophilic bacteria (39, 40). The resulting *P. furiosus* CelB mutants E372D and E372Q were produced in *E. coli* in slightly lower amounts than the wild-type CelB (data not shown). Comparison of the specific activities in the cell extract at 90°C of the mutants E372D and E372Q with that of the wild type showed decreased specific activities of 200- and more than 1,000-fold, respectively.

DISCUSSION

In order to develop a model system for the structure-function analysis of the hyperthermostable *P. furiosus* β-glucosidase, the *celB* gene for this archaeal enzyme was characterized, overexpressed in *E. coli*, and subjected to site-directed mutagenesis. In addition, the transcription and translation initiation signals of the *celB* gene were identified in *P. furiosus*.

Recently, the regulatory regions of various genes from hyperthermophilic members of the *Archaea* have been aligned and used to define consensus sequences (6). Upstream of these genes, the consensus hexanucleotide motif $TTTA^{A}/_{T}A$ was found that resembles the box A positioned −25 to −27 of the transcription initiation site of other archaeal genes (2, 38). This sequence was found to interact with the RNA polymerase of the close relative *Pyrococcus woesei* (30). Downstream of the genes for hyperthermophiles, a conserved stretch of at least four successive T residues was found to be conserved around the site of transcription termination (2, 6). The transcription of the *celB* gene is initiated at an adenine residue that is preceded by the hexanucleotide ATTATA at the conserved distance of 25 nucleotides. This box A sequence is somewhat different from the consensus found in hyperthermophilic members of the *Archaea* with respect to the first nucleotide, which is an adenine instead of a thymidine. Transcription termination sequences could be identified downstream of the stop codon of *celB*. Additionally, an inverted repeat is located downstream of the gene, but its functionality in transcription termination remains to be determined.

Comparison of the NH_2-terminal amino acid residues of the purified β-glucosidase with that deduced from the *celB* gene sequence indicates that the initiation methionine is not removed from the mature enzyme in *P. furiosus* by posttranslational modification as found for the glutamate dehydrogenases from *P. furiosus*, *P. woesei*, and *Pyrococcus endeavori* (formerly ES4) (9, 11, 26). Analysis of the codon usage of the *celB* gene showed a preference for the codons without a CG dinucleotide, which is in line with observations done for other pyrococcal genes (9, 11, 23). Remarkably, the bias towards the CG dinucleotide in the *celB* gene is observed only in the 5′ to 3′ direction and not in the opposite direction, suggesting that it affects translation.

The organization of the *celB* gene indicates that it is transcribed as a single unit, which is substantiated by the presence of transcription initiation signals of an additional open reading frame downstream of *celB*, encoding a $tRNA^{Phe}$ that has its own transcription signals (37). The genetic organization of the *P. furiosus celB* gene is clearly different from that of the *Sulfolobus solfataricus lacS* gene encoding another archaeal glycosyl hydrolase. This latter *lacS* gene is preceded by a gene encoding a putative permease that is oriented in the same direction (29), while *celB* is preceded by a gene that is transcribed in the opposite orientation and does not show the characteristics of a hydrophobic permease (37).

The alignment of the deduced primary sequence of *celB* from *P. furiosus* with other members of glycosyl hydrolase family 1 shows a large number of amino acids that are conserved. Three subgroups can be distinguished within this family: one group consists of the archaeal glycosyl hydrolases with β-glycosidase from *P. furiosus* and those from the extreme thermophilic archaeon *Sulfolobus solfataricus*, a second group includes the glycosidases from the *Eucarya* and *Bacteria*, and the third group comprises the 6-phosphoglycosidases from bacterial origin (Fig. 3). This subdivision is supported by differences in biochemical characteristics of the archaeal glycosidases and those of the other glycosidases. The archaeal glycosidases contain, unlike most other glycosidases, no essential −SH groups for catalysis and form tetramers instead of monomers (21). One other multimeric member of the family is the bacterial glucosidase from the hyperthermophile *Thermotoga maritima*, which is a dimeric enzyme (14). This multimerization in enzymes from extreme thermophiles may be a way to protect groups that are susceptible to modification at high temperatures or to prevent local unfolding and thereby increasing their stability. The increased thermostability of archaeal glycosidases in comparison to other glycosyl hydrolases of this family could be related to the presence of one or more insertions in the archaeal enzymes.

General trends in changes of amino acid composition related to increased stability in proteins from hyperthermophiles compared to their mesophilic homologs include a decreased number of cysteines, an increased overall hydrophobicity, and a decreased average chain flexibility (18). In addition to these general trends, some other differences in amino acid composition are noted when the CelB sequence is compared to that of other glycosyl hydrolases. These include a significantly increased number of acid residues, notably glutamate (42 in *P. furiosus* and on average 33 in other glycosidases). One region within the CelB amino acid sequence from residue 273 to residue 288 shows a high number of charged residues (mainly glutamate) and could be involved in the electrostatic interactions, possibly between different subunits. Whether the local or overall increased number of charged residues that could form

salt bridges is indeed related to the extreme stability of CelB remains to be determined.

The expression of the *P. furiosus celB* gene in *E. coli* was significantly improved by its positioning under control of the *tac* promoter in the pTTQ derivative pLUW510. In this way we were able to control *celB* expression and reach a production level up to 20% of total *E. coli* protein. Successful expression of *P. furiosus* genes for glyceraldehyde-3-phosphate dehydrogenase, DNA polymerase, and α-amylase in *E. coli* has been demonstrated previously, but enzyme production levels were low (23, 36, 41). However, high expression of the *P. furiosus gdh* gene was recently obtained that resulted in functional glutamate dehydrogenase (8, 11, 25). A heat treatment was needed to obtain a fully active enzyme after expression of pyrococcal *gdh* gene in *E. coli* (8). In contrast, such a heat treatment was not required to activate CelB produced in *E. coli*.

Residues that are important for the catalytic activity of glycosidases have been identified by protein engineering (35). Comparison of the CelB produced in *E. coli* with the native enzyme purified from *P. furiosus* showed no significant differences in kinetic and stability properties. Moreover, the NH_2-terminal residues of both enzymes are identical and not post-translationally modified. The observed identity between the *E. coli*- and *P. furiosus*- produced β-glucosidases validates the use of the heterologous enzyme as model for protein engineering studies. This approach was used to identify the active site nucleophile of the *P. furiosus* β-glucosidase involved in catalysis at high temperature. Alignment of glycosyl hydrolases of family 1 shows the conservation of glutamate 372 in the *P. furiosus* β-glucosidase that in the corresponding proteins from the mesophiles *Agrobacterium* sp. and *S. aureus* has been identified as the active site nucleophile at low temperatures (39, 40). Replacement of the glutamate 372 by an aspartate or glutamine residue in the *P. furiosus* β-glucosidase reduced its activity more than 200-fold. The large effect of these relative small changes, i.e., size reduction of the side chain or elimination of the negative charge, indicates that glutamate 372 is essential for catalysis of the pyrococcal CelB. This strongly suggests that this residue is the active site nucleophile involved in catalysis above 100°C and shows the feasibility of protein engineering of this highly stable hydrolase.

ACKNOWLEDGMENTS

The assistance of Fiona Kaper, Peter Steenbakker, Frank Wagner, and Han Zendman is early phases of this work is greatly appreciated. We thank Servé Kengen for helpful discussion.

Part of this work was supported by contract BIOT-CT93-0274 of the European Union.

REFERENCES

1. **Badr, H. R., K. A. Sims, and M. W. W. Adams.** 1994. Purification and characterization of sucrose α-glucohydrolase (invertase) from the hyperthermophilic archaeon *Pyrococcus furiosus*. Syst. Appl. Microbiol. **17**:1–6.
2. **Brown, J. W., C. J. Daniels, and J. N. Reeve.** 1989. Gene structure, organization, and expression in Archaebacteria. Crit. Rev. Microbiol. **16**:287–338.
3. **Casadaban, M. J., J. Chou, and S. N. Cohen.** 1980. In vitro gene fusions that join an enzymatically active β-galactosidase segment to amino-terminal fragments of exogenous proteins: *Escherichia coli* plasmid vectors for the detection and cloning of translational initiation signals. J. Bacteriol. **143**:971–980.
4. **Chirgwin, J. M., A. E. Przybyla, R. J. MacDonald, and W. J. Rutter.** 1979. Isolation of biologically active ribonucleic acid from sources enriched in ribonuclease. Biochemistry **18**:5294–5299.
5. **Constantino, H. R., S. H. Brown, and R. M. Kelly.** 1990. Purification and characterization of an α-glucosidase from a hyperthermophilic archaebacterium, *Pyrococcus furiosus*, exhibiting a temperature optimum of 105 to 115°C. J. Bacteriol. **172**:3654–3660.
6. **Dalgaard, J. Z., and R. A. Garrett.** 1993. Archaeal hyperthermophile genes, p. 535–563. *In* M. Kates et al. (ed.), The biochemistry of Archaea (Archaebacteria). Elsevier, Amsterdam.
7. **Devereux, J., P. Haeberli, and O. Smithies.** 1990. A comprehensive set of sequence analysis programs for the VAX. Nucleic Acids Res. **12**:387–395.
8. **Diruggiero, J., and F. T. Robb.** 1995. Expression and in vitro assembly of recombinant glutamate dehydrogenase from the hyperthermophilic archaeon *Pyrococcus furiosus*. Appl. Environ. Microbiol. **61**:159–164.
9. **Diruggiero, J., F. T. Robb, R. Jagus, H. H. Klump, K. M. Borges, M. Kessel, X. H. Mai, and M. W. W. Adams.** 1993. Characterization, cloning, and in vitro expression of the extremely thermostable glutamate dehydrogenase from the hyperthermophilic archaeon, ES4. J. Biol. Chem. **268**:17767–17774.
10. **Eggen, R. I. L., A. C. M. Geerling, M. S. M. Jetten, and W. M. de Vos.** 1991. Cloning, expression and sequence analysis of the genes for carbon monoxide dehydrogenase of *Methanothrix soehngenii*. J. Biol. Chem. **266**:6883–6887.
11. **Eggen, R. I. L., A. C. M. Geerling, W. G. B. Voorhorst, R. Kort, and W. M. de Vos.** 1994. Molecular and comparative analysis of the hyperthermostable *Pyrococcus furiosus* glutamate dehydrogenase and its gene. Biocatalysis **8**:131–141.
12. **Fiala, G., and K. O. Stetter.** 1986. *Pyrococcus furiosus* sp. nov. represents a novel genus of marine heterotrophic archaebacteria growing optimally at 100°C. Arch. Microbiol. **145**:56–61.
13. **Fontana, A.** 1991. How nature engineers proteins thermostability, p. 89–113. *In* G. de Presco (ed.), Life under extreme conditions: biochemical adaptation. Springer, Berlin.
14. **Gabelsberger, J., W. Liebl, and K.-H. Schleifer.** 1993. Purification and properties of recombinant β-glucosidase of the hyperthermophilic bacterium *Thermotoga maritima*. Appl. Microbiol. Biotechnol. **40**:44–52.
15. **Gibson, T. J.** 1984. Studies on the Epstein-Barr-Virus genome. Ph.D. thesis. Cambridge University, Cambridge.
16. **Henrissat, B.** 1991. A classification of glycosyl hydrolases based on amino acid sequence similarities. Biochem. J. **280**:309–316.
17. **Henrissat, B., and A. Bairoch.** 1993. New families in the classification of glycosyl hydrolases based on amino acid sequence similarities. Biochem. J. **193**:781–788.
18. **Hensel, R.** 1993. Proteins of extreme thermophiles, p. 209–221. *In* M. Kates et al. (ed.), The biochemistry of Archaea (Archaebacteria). Elsevier, Amsterdam.
19. **Kengen, S. W. M., F. A. M. de Bok, N. D. van Loo, C. Dijkema, A. J. M. Stams, and W. M. de Vos.** 1994. Evidence for the operation of a novel Embden-Meyerhof pathway that involves ADP-dependent kinases during sugar fermentation by *Pyrococcus furiosus*. J. Biol. Chem. **269**:17537–17541.
20. **Kengen, S. W. M., E. J. Luesink, A. J. M. Stams, and A. J. B. Zehnder.** 1993. Purification and characterization of an extremely stable β-glucosidase from the hyperthermophilic archaeon *Pyrococcus furiosus*. Eur. J. Biochem. **213**:305–312.
21. **Kengen, S. W. M., and A. J. M. Stams.** 1994. An extremely thermostable β-glucosidase from the hyperthermophilic archaeon *Pyrococcus furiosus*: a comparison with other glycosidases. Biocatalysis **11**:79–88.
22. **Kengen, S. W. M., and A. J. M. Stams.** 1994. Formation of L-alanine as a reduced end product in carbohydrate fermentation by the hyperthermophilic archaeon *Pyrococcus furiosus*. Arch. Microbiol. **161**:168–175.
23. **Laderman, K. A., K. Asada, T. Uemori, H. Mukai, Y. Taguchi, I. Kato, and C. B. Anfinsen.** 1993. α-Amylase from the hyperthermophilic archaebacterium *Pyrococcus furiosus*—cloning and sequencing of the gene and expression in *Escherichia coli*. J. Biol. Chem. **268**:24402–24407.
24. **Laemmli, U. K.** 1970. Cleavage of structural proteins during the assembly of the head of bacteriophage T4. Nature (London) **227**:680–685.
25. **Lebbink, J. H. G., R. I. L. Eggen, A. C. M. Geerling, V. Consalvi, R. Chiaraluce, R. Scandurra, and W. M. de Vos.** Exchange of domains of glutamate dehydrogenase from the hyperthermophilic archaeon *Pyrococcus furiosus* and the mesophilic bacterium *Clostridium difficile*: effects on catalysis, thermoactivity and stability. Prot. Eng., in press.
26. **Maras, B., S. Valiante, R. Chiaraluce, V. Consalvi, L. Politi, M. Derosa, F. Bossa, R. Scandurra, and D. Barra.** 1994. The amino acid sequence of glutamate dehydrogenase from *Pyrococcus furiosus*, a hyperthermophilic archaebacterium. J. Prot. Chem. **13**:253–259.
27. **Mukund, S., and M. W. W. Adams.** 1991. The novel tungsten-iron-sulfur protein of the hyperthermophilic archaebacterium, *Pyrococcus furiosus*, is an aldehyde ferredoxin oxidoreductase—evidence for its participation in a unique glycolytic pathway. J. Biol. Chem. **266**:14208–14216.
28. **Owen, J. L., C. Hay, D. M. Schuster, and A. Rashtchian.** 1994. A highly efficient method of site-directed mutagenesis using PCR and UDG cloning. Focus **16**:39–44.
29. **Prisco, A., M. Moracci, M. Rossi, and M. Ciaramella.** 1995. A gene encoding a putative membrane protein homologous to the major facilitator superfamily of transporters maps upstream of the β-glycosidase gene in the archaeon *Sulfolobus solfataricus*. J. Bacteriol. **177**:1614–1619.
30. **Rowlands, T., P. Baumann, and S. P. Jackson.** 1994. The TATA-binding protein: a general transcription factor in eukaryotes and archaebacteria. Science **264**:1326–1329.
31. **Sambrook, J., E. F. Fritsch, and T. Maniatis.** 1989. Molecular cloning: a laboratory manual, 2nd ed. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y.
32. **Sanger, F., S. Nicklen, and A. R. Coulson.** 1977. DNA sequencing with chain-terminating inhibitors. Proc. Natl. Acad. Sci. USA **74**:5463–5467.

33. **Schäfer, T., and P. Schonheit.** 1991. Pyruvate metabolism of the hyperthermophilic archaebacterium *Pyrococcus furiosus*—acetate formation from acetyl-CoA and ATP synthesis are catalyzed by an acetyl-CoA synthetase (ADP forming). Arch. Microbiol. **155**:366–377.
34. **Studier, F. W., B. A. Moffat, and J. Dunn.** 1986. Use of bacteriophage T7 RNA polymerase to direct high level expression of cloned genes. J. Mol. Biol. **189**:113–130.
35. **Svensson, B., and M. Sogaard.** 1993. Mutational analysis of glycosylase function. J. Biotechnol. **29**:1–37.
36. **Uemori, T., Y. Ishino, H. Toh, K. Asada, and I. Kato.** 1993. Organization and nucleotide sequence of the DNA polymerase gene from the archaeon *Pyrococcus furiosus*. Nucleic Acids Res. **21**:259–265.
37. **Voorhorst, W. G. B., and W. M. de Vos.** Unpublished results.
38. **Wettach, J., H. P. Gohl, H. Tschochner, and M. Thomm.** 1995. Functional interaction of yeast and human TATA-binding proteins with an archaeal RNA polymerase and promoter. Proc. Natl. Acad. Sci. USA **92**:472–476.
39. **Whiters, S. G., K. Rupitz, D. Trimbur, and R. A. J. Warren.** 1992. Mechanistic consequences of mutation of the active site nucleophile Glu 358 in *Agrobacterium* β-glucosidase. Biochemistry **31**:9979–9985.
40. **Witt, E., R. Frank, and W. Hengstenberg.** 1993. 6-Phospho-β-galactodisases of Gram-positive and 6-phospho-β-galactosidase B of Gram-negative bacteria: comparison of structure and function by kinetic and immunological methods and mutagenesis of the *lacG* gene of *Staphylococcus aureus*. Prot. Eng. **6**:913–920.
41. **Zwickl, P., S. Fabry, C. Bogedain, A. Haas, and R. Hensel.** 1990. Glyceraldehyde-3-phosphate dehydrogenase from the hyperthermophilic archaebacterium *Pyrococcus woesei*: characterization of the enzyme, cloning and sequencing of the gene, and expression. J. Bacteriol. **172**:4329–4338.

Chapter 3

Genetic and Biochemical Characterization of a Short-Chain and an Iron-Containing Alcohol Dehydrogenase from the Hyperthermophilic Archaeon *Pyrococcus furiosus*

Wilfried G.B. Voorhorst, Vincent Wittenhorst, Yannick Gueguen,
John van der Oost, and Willem M. de Vos

submitted for publication

Abstract

Analysis of the genomic region preceding the *celB* gene of the hyperthermophilic archaeon *Pyrococcus furiosus*, revealed the presence of two tandem genes, oppositely orientated of *celB*. These have been designated *adhA* and *adhB*, since their predicted products showed high homology to short-chain and iron-containing NADP(H)-dependent alcohol dehydrogenases, respectively. Overexpression of the *adhA* and *adhB* genes in *Escherichia coli* showed the presence of two prominent, heat-stable proteins with an apparent subunit size of 27 kDa and 44 kDa, respectively, that corresponded closely to that calculated from their gene sequence. AdhB showed a low, but reproducible NADP(H)-dependent alcohol dehydrogenase activity with methanol as substrate, that was rapidly lost, preventing further characterization. In contrast, AdhA showed a stable NADP(H)-dependent alcohol dehydrogenase activity and was purified to homogeneity from the overproducing *E.coli* cells. Activity measurements and immunoblot analysis of cell-extracts showed that the *adhA* gene was expressed in *P.furiosus* and resulted in an AdhA with identical subunit size as that found in the overproducing *E.coli*. The purified AdhA showed an apparent molecular mass of 95 kDa indicating a homotetrameric structure and appeared to have an optimal enzymatic activity at 90°C at which temperature it had a half-life of 22.5 h. AdhA reversibly converted alcohols to the corresponding aldehydes, with a preference for the oxidation of secondary over the primary alcohols. Highest conversion rates were found towards 2-pentanol at pH 10.5 in the oxidation reaction and towards pyruvaldehyde in the reduction reaction at neutral pH, suggesting that aldehyde reduction is the physiological role of AdhA in *P.furiosus*.

Introduction

The alcohol dehydrogenases (ADHs) that are dependent on NAD or NADP for catalysis, can be divided into three groups; (i) long-chain or zinc-containing ADH, (ii) short-chain or insect-type ADH, and (iii) iron-containing ADH (17). Representatives of all three groups have been extensively studied in *Bacteria* and *Eucarya* on the biochemical and genetic level with specific attention for their substrate specificity, metabolic functions and biotechnological potential (2, 16, 17).

Within the *Archaea*, the third domain of life, so far only a limited number of ADHs have been characterized. In some methanogens NAD(P)-dependent ADHs have been found that are involved in utilization of alcohols, other than methanol, as substrate (3). Instead of utilizing alcohols, several hyperthermophilic archaea that grow optimally above 80°C are known to produce alcohols as fermentation product (20). Alcohol dehydrogenases identified in these extreme microorganisms include a NADP-dependent ADH in *Hyperthermus butylicus*, that preferably converts secondary alcohols (2), and three iron-containing ADHs from *Thermococcus* sp. that have been purified and showed to be dependent on NADP(H) for catalysis (13, 14, 15). For two of these thermococcal enzymes a role has been suggested in the dispose of reduction equivalents through the formation of alcohols (14, 15). The genetic characterization of archaeal ADHs is limited to the gene coding for an iron-containing ADH from *Thermococcus* strain AN1 (13) and two closely related zinc-containing ADHs from two *Sulfolobus* spp. for which the genes have recently been cloned and expressed in *Escherichia coli* (1). There is considerable interest in ADHs from hyperthermophiles, since they are potential alternatives for the less thermostally and chemically stable counterparts that are currently used as industrial biocatalysts (2).

The most extensively studied representative of the hyperthermophilic archaea is *Pyrococcus furiosus* that is able to utilize carbohydrates and polypeptides (5). The main fermentation products include H_2 or H_2S, CO_2, acetate, and alanine (5, 8, 10). In addition, the production of small amounts of ethanol has been documented (8). Here we report the identification and characterization of two tandem genes, *adhA* and *adhB*, encoding different types of ADHs from *P.furiosus* and their expression in *E. coli*. One of these, AdhA, that showed high homology with short-chain ADHs has been purified to homogeneity and characterized with respect to substrate specificity, stereoselectivity, kinetics and stability.

Materials and methods

Strains, plasmids and cloning procedure. *P.furiosus* (DSM3638) was cultured at 98°C in synthetic seawater as previously described using cellobiose (10 mM) as growth substrate (9). All plasmids used are derivatives of pTZ19R (Pharmacia Biotechnologies, Upsalla, Sweden) containing chromosomal DNA fragments of *P.furiosus* (Fig. 1) and were constructed in *E.coli* TG1 (6) using established methods

(19). Plasmids pLUW500 and pLUW501 have been described previously (22). The plasmids pLUW521 and pLUW520 were constructed by reducing the upstream region of the *celB* gene in pLUW500 by deletion of either the *Sma*I-*Sma*I fragment or the *Eco*RV-*Sma*I fragment, respectively (Fig.1), using the vector located *Sma*I-site.

Nucleotide sequence analysis. Automated nucleotide sequence analysis of plasmid DNA was carried out on an Applied Biosystems 373A sequencer (Applied Biosystems Inc. Foster City, CA) using a Prism™ Ready Reaction DyeDeoxy™ Terminator cycle sequencing kit and a Li-Cor 4000L sequencer (Li-Cor Inc. Lincoln, NE) using the Thermo Sequenase cycle sequencing kit with 7-deaza-dGTP (Amersham International plc, UK) and IRD41-labelled oligonucleotides (MWG-Biotech, München, Germany).

Computer analysis of nucleotide and deduced amino acid sequences were carried out with the PC/GENE program version 5.01 (IntelliGenetics Inc.) and the GCG package version 7.0 (4) at the CAOS/CAMM centre of the University of Nijmegen (The Netherlands). Amino acid sequence were aligned using Clustal W (1.6) and from these results phylogenetic trees were constructed by the neighbour-joining method (18) using the TREECON program (21).

Analysis of expression products and purification of ADH. Cells of *E.coli* TG1 harboring pTZ19R-derivatives were grown overnight in 5 ml L-broth with 100 μg of ampicillin per ml, harvested by centrifugation. Cells were resuspended in 0.2 ml Tris-buffer (20 mM Tris-HCl, pH 7.8). For small scale experiments (up to 10 ml of culture), cells were disrupted by sonication and the cell extracts were incubated for 15 min at 70°C. Cell debris and denatured *E.coli* proteins were removed by centrifugation and the supernatants were subjected to further analysis. For large scale experiments (1 liter of culture) aimed at the purification of ADH, cells were disrupted using a French press. RNase (50 mg/ml), DNase (50 mg/ml) and $MgCl_2$ (1 mM) were added, and the mixture was incubated for 10 min at 37°C followed by the denaturation step at 70°C. After centrifugation, the supernatant was two-fold diluted with Tris-buffer and loaded onto a Q-Sepharose High Performance (Pharmacia) column (5.0 by 15.0 cm) that was preequilibrated in the same Tris-buffer. Bound proteins were eluted by a linear gradient of sodium chloride (0-1 M in Tris-buffer). Active fractions were pooled and concentrated by ultrafiltration with a stirred cell using a membrane with a nominal cut-off of 10 kDa (Amicon, Inc., Beverly, MA).

Polyacrylamide gel electrophoresis and Western blotting. Protein composition was analyzed by sodium dodecyl sulphate-polyacrylamide (11%) gel electrophoresis (SDS-PAGE) (11). Protein samples for SDS-PAGE were prepared by heating for 5 min at 100°C in an equal volume of sample buffer (0.1 M Tris-HCl, 5% SDS, 0.9% 2-mercaptoethanol, 20% glycerol, pH 6.8).

For Western blot analysis proteins were separated by SDS-PAGE and transferred to a nitrocellulose membrane by electroblotting using a semi-dry transfer cell (Bio-Rad Laboratories Hercules, CA). Purified AdhA was used to raise rabbit antibodies, and crude serum was subsequently used in immunoblot experiments. Detection of AdhA on nitrocellulose membranes was performed using a goat anti-rabbit alkaline phosphatase conjugate as described in the Protoblot Immunoscreening system protocol (Promega Corporation Madison, WI).

Activity assays. Alcohol oxidation and aldehyde reduction were determined at 70°C, unless stated otherwise, by following either the reduction of NADP or the oxidation of NADPH spectrophotometrically at 340 nm using a Beckman DU7500 spectrophotometer with supplied temperature controller (Beckman Instruments, Inc. Fullerton, CA). Each oxidation reaction contained 50 mM glycine (pH 10.5), 0.25 M of alcohol and 0.28 mM NADP. The aldehyde reduction contained 0.1 M sodium phosphate (pH 7.0), 0.34 M aldehyde and 0.28 mM NADPH. In addition these reactions were performed with the non-phosphorylated cofactors NAD and NADH. One unit of ADH was defined as the formation or removal of 1 mmol of NADPH per min. In all assays the reaction was initiated by addition of enzyme. Protein concentration was determined using Bradford reagents (Bio-Rad) with bovine serum albumin as a standard.

Characterization of AdhA. The optimum pH for alcohol oxidation was determined in the pH-range of 6.3 to 11.5 using a 50 mM glycine buffer, whereas the optimum pH for aldehyde reduction was determined in the pH-range of 5.0 to 7.0 using a buffer containing 100 mM citrate and 100 mM sodium phosphate. The kinetic parameters K_m and V_{max} were calculated from multiple measurements using the program Enzfitter (12). The thermostability of AdhA (0.18 mg/ml) was determined in Tris-buffer by measuring the residual activity after various times of incubation at 80°, 90°C, or 100°C.

Nucleotide sequence accession number. The nucleotide sequence reported has been submitted to GenBank/EMBL Data Bank with the accession number AF013603

Results

Characterization of *adhA* and *adhB* genes and their deduced proteins. Nucleotide sequence analysis of the 7.7-kb *P.furiosus* DNA in plasmid pLUW501 (22), revealed the presence of four clustered complete open reading frames (ORFs) in opposite orientation of the previously described *celB* gene, encoding a ß-glucosidase (Fig.1) (22). The last two ORFs will be discussed elsewhere and include the *lamA* gene coding for a secreted ß-1,3-endoglucanase (7) and the *birA* gene probably coding for a biotin ligase (23). The first ORF has been designated *adhA* (see below) and contains an ATG start codon preceded by a purine-rich sequence, that is likely to function as a ribosome-binding site. The stop codon TGA of *adhA* was immediately followed by another purine-rich stretch of 10 nucleotides preceding an ATG start codon that initiates a second gene, designated *adhB* (see below) (Fig.1).

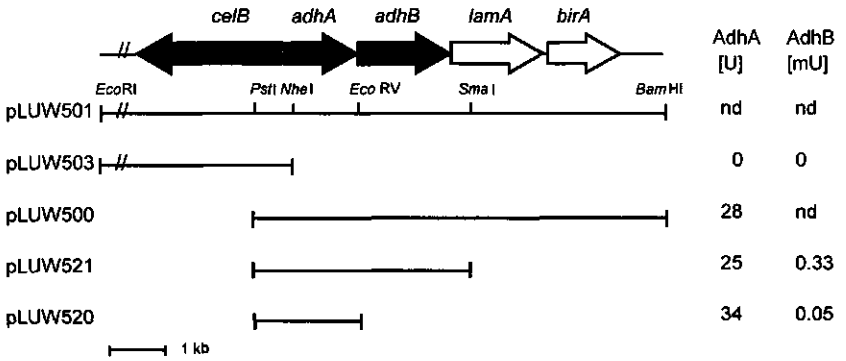


Figure 1. Schematic representation of the organization of the *P.furiosus celB* locus. The location and direction of the genes deduced from the DNA sequence are shown as is a restriction map with relevant sites used for plasmid constructions. The *lamA* encodes a secreted ß-1,3-endoglucanase (7) and *birA* a biotin ligase-like protein (23). The right panel shows the specific activity of AdhA and AdhB in extracts of *E.coli* harboring the indicated plasmids, which were determined using the substrates 2-pentanol or methanol, respectively; nd: not determined.

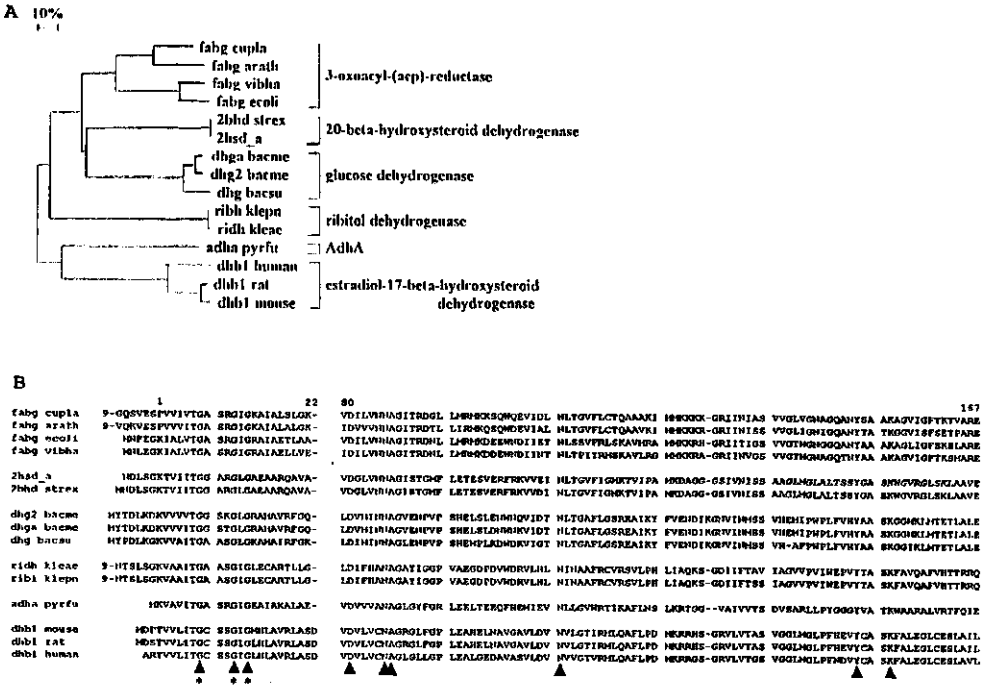


Figure 2. The phylogenetic relation (A) and partial alignment (B) of the *P.furiosus* AdhA and members of short-chain ADHs. The dendrogram shows the phylogenetic relationship of the *P.furiosus* AdhA (adha pyrfu) with members of the same group that have an identified enzymatic activity. The relevant parts of the amino acid alignment, used for the generation of the phylogenetic tree is shown: the numbering at the top corresponds with that of the *P.furiosus* AdhA. Abbreviations correspond with those deposited in the Swiss Prot and Genbank database or were deduced using the conventional procedures. Arrows indicate highly conserved residues within the complete group of short-chain ADHs; those marked with stars indicate residues that may interact with the cofactor.

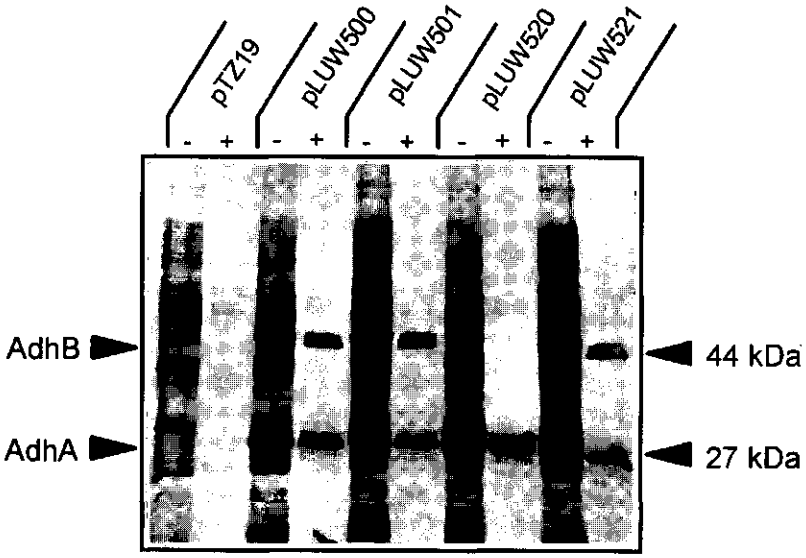


Figure 3. SDS-PAGE of cell-extracts of plasmid-harboring *E.coli* TG1. Cell-extracts of *E.coli* harboring the indicated plasmids either prior (-) or after heat-treatment (+) were separated by SDS-PAGE and the proteins were stained with Coomassie Brilliant Blue.

If the assumed initiation codons are used, *adhA* and *adhB* genes could encode proteins with a size of 233 and 389 amino acids and calculated molecular masses of 26 kDa and 43 kDa, respectively. The deduced amino acid sequence of AdhA from *P.furiosus* showed high homologies with members of the group of short-chain ADHs (Fig. 2). The highest homology was found with eucaryal estradiol 20-ß-hydroxysteroid dehydrogenases (35 % identity). The deduced amino acid sequence of the pyrococcal AdhB showed high homologies with members of the group of iron-containing ADHs and up to 35% identity was observed with the two closely related NADH-dependent butanol dehydrogenases from *Clostridium acetobutylicum* and 34% identity with the only other archaeal member of this group, *Thermococcus* strain AN1 (13).

Cloning and overexpression of *adhA* and *adhB* genes in *E.coli*. In cell-extracts of *E.coli* harboring pLUW501 two prominent heat-stable proteins could be detected by SDS-PAGE with apparent molecular masses of 27 kDa and 44 kDa (Fig. 3). Together, these proteins accounted for approximately 10% of the soluble protein

fraction of the cell-extract (Fig.3). Deletion of the *celB* gene resulting in pLUW500 did not affect the presence of these proteins in heat-treated cell-extracts of *E.coli* harboring pLUW500, but they were not found in *E.coli* harboring pLUW503, containing only the *celB* gene (Fig. 1; data not shown). Further deletion studies resulted in pLUW521, which upon introduction in *E.coli* gave rise to the production of both heat-stable proteins, and pLUW520 that produced only the 27 kDa protein (Figs. 1 and 3). These results indicate that the 27-kDa protein is encoded by the *adhA* gene and the 44-kDa protein is the product of the *adhB* gene.

The *adhA* and *adhB* genes code for different ADH activities. In heat-treated cell-extracts of *E.coli* overproducing the pyrococcal AdhA and AdhB high ADH activities were found at 70 °C that were dependent on the presence of NADP or NADPH (Fig. 1). To distinguish between the activities of the two pyrococcal ADHs, use was made of the observation that purified AdhA showed low dehydrogenase activity with methanol as substrate, but high activity with 2-pentanol (see below). Extracts of *E.coli* harboring *adhA*-expressing plasmids, including pLUW520 carrying only the *adhA* gene, showed significant ADH activity with 2-pentanol as substrate (Fig.1). Extracts of *E.coli* harboring *adhB*-expressing plasmids, showed a significant but low ADH specific activity with methanol as substrate that was substantially higher than that of *E.coli* harboring pLUW520 (Fig. 1). The latter can be attributed to the side activity of AdhA towards the substrate methanol (see below). These findings indicate that both *adhA* and *adhB* code for functional ADHs. However, the methanol-dependent ADH activity was rapidly lost (80 % during 1 day of storage at 4°C) and could not be restored by the addition of divalent cations, such as iron, zinc or magnesium, preventing a more detailed characterization.

Purification, characterization, and substrate specificity of AdhA. The pyrococcal AdhA was purified to homogeneity from a heat-treated cell-extract of *E.coli* harboring pLUW520, by anion-exchange and size-exclusion chromatography, confirming a molecular subunit mass of 27 kDa (see below). The molecular mass of the native AdhA was estimated to be 95 kDa by size-exclusion chromatography, which indicated a tetrameric structure.

The substrate specificity of AdhA in the oxidation reaction was analyzed using a range of alcohols, including primary alcohols (C_1 to C_{11}), secondary alcohols (C_3 to C_8) (Fig.4), and alcohols containing more than one hydroxyl-group (Table 1). The highest

activity of AdhA (41.6 U/mg) was found when 2-pentanol was oxidized. A broad range of primary alcohols could be oxidized by AdhA and include substrates with 2-10 carbon atoms (Fig.4). Similarly, AdhA showed activity towards a broad range of secondary alcohols that on average were faster oxidized than the primary alcohols (Fig.4). Alcohols with a second or a third hydroxyl-group were less efficiently converted by AdhA, most likely due to steric hindrance (Table 1).

The highest reduction rate of AdhA was observed with pyruvaldehyde. The other aldehydes tested, showed only 10% or less of this activity (Table 2).

Secondary alcohols can have two stereoisomers that may show different enzymatic conversion rates. The stereoselectivity of AdhA was tested using the two enantiomers of 2-butanol and revealed only a slight preference for the S-2-butanol (21.8 U/mg and 32.3 U/mg with R- and S-2-butanol, respectively).

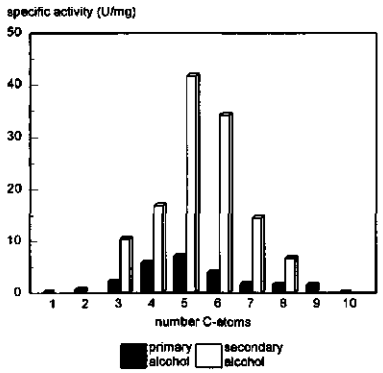


Figure 4. Specific activities of the purified *P.furiosus* AdhA towards primary and secondary alcohols

Optimum, kinetics and stability of AdhA. The oxidation reaction catalyzed by AdhA showed a pH optimum between 10 and 11. The aldehyde reduction by AdhA showed maximal activity in a broad pH range with an optimum at pH 7.5 and activities that were only 20 % lower at pH 6.7 and pH 9.0. The optimum activity of AdhA was at 90°C, but due to instability of the cofactors at that temperature all other activity measurements were performed at 70°C. Kinetic properties of AdhA were determined for both the oxidation and the reduction reaction using the substrate that showed the highest activity, 2-pentanol and pyruvaldehyde, respectively (Table 3). The results indicate that AdhA has the highest affinity for pyruvaldehyde and is dependent on the phosphorylated cofactor for catalysis.

Table 1. Substrate specificity of the purified *P.furiosus* AdhA towards substituted alcohols.

Substrate	Specific activity (U/mg)
ethanol	0.7
ethanediol	1.0
1-propanol	2.1
2-propanol	10.4
1,2-propanediol	2.9
glycerol	0.2
2-methyl-1-propanol	6.2
2-methyl-2-propanol	0.2
1-butanol	5.1
2-butanol	16.8
1,2-butanol	2.4
1,3-butanol	6.9
2,3-butanol	5.5

Table 2. Substrate specificity of the purified *P.furiosus* AdhA towards aldehydes.

Substrate	Specific activity U/mg)
pyruvaldehyde	30.2
acetoin	3.2
propionaldehyde	2.6
glyceraldehyde	2.6
pyruvate	1.4
propionate	0.7
acetone	0.2
hydroxyacetone	1.3
dihydroxyacetone	<0.05

Table 3. Kinetic parameters of the purified *P.furiosus* AdhA in the reduction or oxydation of the substrates pyruvate aldehyde or 2-pentanol, respectively.

	V_{max} (U/mg)	K_m (mM)
pyruvaldehyde	32.3 ± 2.2	3.3 ± 0.2
NADPH		0.08 ± 0.008
2-pentanol	41.6 ± 2.7	60.8 ± 2.9
NADP		0.08 ± 0.009

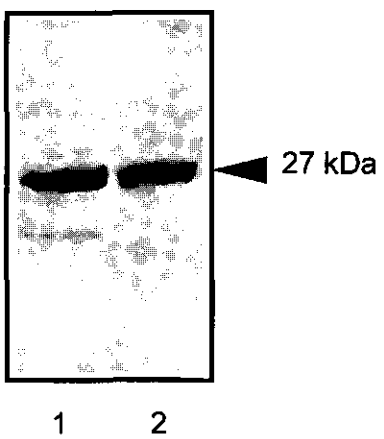


Figure 5. Immunoblot analysis of AdhA in *P.furiosus*. A cell-extract of *P.furiosus* (lane 1) (3 mg) and AdhA purified from *E.coli* (lane 2) (0.03 mg) were separated by SDS-PAGE, blotted on a nitrocellulose membrane and probed with AdhA antiserum. The apparent molecular sizes in kilodaltons are indicated on the left.

Storage of AdhA for several months at 4°C did not affect the enzyme activity significantly. The resistance to thermal inactivation of AdhA at higher temperatures was analyzed and half-life values of 150 h, 22.5 h and 25 min were found at temperatures of 80°C, 90°C and 100°C, respectively.

The presence of ADH activity in *P.furiosus*. Activity measurements of AdhA were performed in cell-extracts of *P.furiosus* using 2-pentanol as substrate and showed a slow, but significant oxidation rate of this substrate (0.27 U/mg), which would correspond with approximately 0.7 % of the total soluble protein fraction. To further confirm the presence of AdhA in *P.furiosus*, immunoblot experiments were performed using antibodies raised against the AdhA purified form *E.coli* (Fig. 5). The results show that AdhA is produced in *P.furiosus* grown on cellobiose and confirm that its size on SDS-PAGE is identical to that of the AdhA produced by *E.coli*.

Discussion

In *P.furiosus* two tandem ADH genes, designated *adhA* and *adhB*, have been identified that were positioned adjacent and in opposite orientation of the *celB* gene. *E.coli* harboring plasmids that contained these two genes overproduced (i) AdhB, an iron-containing ADH, that rapidly lost its activity, and (ii) AdhA, a short-chain ADH, that was characterized in detail. AdhA was found to be produced in an active form in *P.furiosus* as could be concluded from both activity measurements and immunological detection. Overexpression of the *adhA* and *adhB* genes in *E.coli* was not controlled by the IPTG-inducable vector-located *lac*-promoter indicating that the region upstream of the genes is involved in transcription initiation, as was previously found for the *celB* gene (22).

The deduced primary sequence of AdhA from *P.furiosus* showed significant homology with a number of short-chain ADHs, but to the best of our knowledge it represents the first described archaeal member of this ADH family. Like all other short-chain ADHs characterized to date, the pyrococcal AdhA contains the predicted coenzyme-binding site at the N-terminus surrounded by highly conserved residues (Fig. 2) (16, 17). Alignment of the deduced amino acid sequence of the pyrococcal AdhA with other short-chain ADHs showed that all other known features of short-chain ADHs, including Asp59, Lys184, and Tyr180 as ionizable group, are conserved in this

archaeal AdhA (Fig. 2) (16, 17). Additionally, the conservation of a large number of glycine residues (9 of 19 residues) was noted, indicating that the conformational conservation found for short-chain ADHs from bacterial and eucaryal origin can be extended to the domain of *Archaea* (16).

The deduced primary sequence of the second ADH from *P.furiosus*, AdhB, shows high homologies with members of the group of iron-containing ADHs, including that of *Thermococcus* AN1. This group is characterized by the size of its primary product of approximately 385 residues, and by the presence of four conserved histidine residues that were also found in AdhB (17). Three of these histidine residues are predicted to be involved in the iron-binding, which appears to be involved in the catalytic activity as has been demonstrated for the iron-containing ADH from *Zymomonas mobilis* (17). Remarkably, both archaeal representatives of the iron-containing ADH family showed an additional proline residue within the putative iron-binding motif the function of which remains to be established. The instability of the pyrococcal AdhB produced by *E.coli* prevented a detailed characterization and comparison with the other archaeal ADHs.

The functional expression of the *P.furiosus adhA* gene in *E.coli* and the stability of its product AdhA allowed for a simple two-step purification of this short-chain ADH and a more detailed biochemical characterization. For the conversion of alcohols to the corresponding aldehyde as well as for the reversed reaction, AdhA could only use the phosphorylated cofactor NADP or NADPH. AdhA showed a broad substrate specificity with primary and secondary alcohols and revealed an optimum for substrates with a carbon chain of 5 residues. A preference was observed for the secondary alcohols that are converted faster than the corresponding primary alcohol.

The pyrococcal AdhA has a half-life value of 150 h at 80°C and is the most thermostable short-chain ADH known to date. In contrast, the stability of the iron-containing AdhB produced by *E.coli* was found to be low as reflected by its rapid loss of activity, which prevented its biochemical characterization. This instability may be explained by the loss iron, as has been described for the other archaeal iron-containing ADH (14, 15). However, attempts to restore the activity by addition of iron or other divalent cations were not successful.

The physiological function of AdhA is most likely the reduction of the aldehyde, since AdhA displayed a higher affinity for the aldehydes than for the alcohols (Table 3). The high conversion rate found with pyruvaldehyde suggests that the natural substrate of AdhA in *P.furiosus* may have a structure similar to this aldehyde. The

conversion of the aldehyde to the corresponding alcohol may be a means to remove reduction equivalents and to regenerate the cofactor. Alternatively, or combined with this role as electron sink, AdhA and possibly AdhB may be involved in the metabolism of ß-linked glucose polymers. This suggestion is based on the localization of the *adhA* and *adhB* genes in the vicinity of the *lamA* and *celB* genes, which are coordinately expressed and whose products are involved in carbohydrate metabolism of ß-linked glucose polymers in *P.furiosus* (7, 22, 23). If so, this would contrast with *Thermococcus* sp. where the iron-containing ADHs are induced during growth on proteins and peptides without elemental sulphur (14). Under these conditions an excess of reduction equivalents may be disposed in the form of alcohols by the reduction of the aldehydes that are produced by different aldehyde ferredoxin oxidoreductases (14).

In conclusion, we have identified two *P.furiosus* genes that encode functional NAD(P)-dependent alcohol dehydrogenases belonging to ADH superfamily and include a short-chain and an iron-containing ADH. The observed conserved primary structure of the pyrococcal AdhA, suggests the presence of a short-chain ADH in the last common ancestor, i.e before the splitting of the phylogenetic tree.

Acknowledgements.

We thank Tony van Kampen and Ans Geerling for assistance with the sequence analysis. Part of this work was supported by contracts BIOT-CT93-0274 and BIOT-CT96-0488 of the European Union.

References

1. **Cannio, R., G. Fiorentino, P. Carpinelli, M. Rossi and S. Bartolucci.** 1996. Cloning and overexpression in *Escherichia coli* of the genes encoding NAD-dependent alcohol dehydrogenase from two *Sulfolobus* species. J. Bacteriol. **178:**301-305.
2. **Cowan, D. A.** 1992. Biotechnology of the Archaea. Trends Biotech. Sci. **10:**315-323.
3. **Daniels, L.** 1993. Biochemistry of metanogenesis, p. p. 41-112. *In* M. Kates *et al.* (ed.), The biochemistry of Archaea (Archaeabacteria), Elsevier, Amsterdam.
4. **Devereux, J., P. Haeberli and O. Smithies.** 1990. A comprehensive set of sequence analysis programs for the VAX. Nucl. Acids Res. **12:**387-395.
5. **Fiala, G. and K. O. Stetter.** 1986. *Pyrococcus furiosus* sp. nov. represents a novel genus of marine heterotrophic archaebacteria growing optimally at 100°C. Arch. Microbiol. **145:**56-61.
6. **Gibson, T. J.** 1984. Studies on the Epstein-Barr-Virus genomeed. Cambridge, UK.

7. **Gueguen, Y., W. G. B. Voorhorst, J. van der Oost and W. M. de Vos.** 1997. Molecular and biochemical characterization of an endo-β-1,3-glucanase of the hyperthermophilic archaeon *Pyrococcus furiosus*. J. Biol. Chem. (in press).
8. **Kengen, S. W. M., F. A. M. de Bok, N. D. van Loo, C. Dijkema, A. J. M. Stams and W. M. de Vos.** 1994. Evidence for the operation of a novel Embden-Meyerhof pathway that involves ADP-dependent kinases during sugar fermentation by *Pyrococcus furiosus*. J. Biol. Chem. **269**:17537-17541.
9. **Kengen, S. W. M., E. J. Luesink, A. J. M. Stams and A. J. B. Zehnder.** 1993. Purification and characterization of an extremely stable β-glucosidase from the hyperthermophilic archaeon *Pyrococcus furiosus*. Eur. J. Biochem. **213**:305-312.
10. **Kengen, S. W. M., A. J. M. Stams and W. M. de Vos.** 1996. Sugar metabolism of hyperthermophiles. FEMS Microbiol. Rev. **18**:119-137.
11. **Laemmli, U. K.** 1970. Cleavage of structural proteins during the assembly of the head of bacteriophage T4. Nature **227**:680-685.
12. **Leatherbarrow, R. J.** (1987). *ENZFITTER*. In Cambridge, UK: Elsevier-Biosoft.
13. **Li, D. and K. J. Stevenson.** 1997. Purification and sequence analysis of a novel NADP(H)-dependent type III alcohol dehydrogenase from *Thermococcus* strain AN1. J.Bacteriol. **179**:4433-4437.
14. **Ma, K., H. Loessner, J. Heider, M. K. Johnson and M. W. W. Adams.** 1995. Effects of elemental sulfur on the metabolism of the deep-sea hyperthermophilic archaeon *Thermococcus* strain ES-1: characterization of a sulfur-regulated, non-heme iron alcohol dehydrogenase. J. Bacteriol. **177**:4748-4756.
15. **Ma, K., F. T. Robb and M. W. W. Adams.** 1994. Purification and characterization of NADP-specific alcohol dehydrogenase and glutamate dehydrogenase from the hyperthermophilic archaeon *Thermococcus litoralis*. Appl. Environ. Microbiol. **60**:562-568.
16. **Persson, B., M. Krook and H. Jornvall.** 1991. Characteristics of short-chain alcohol dehydrogenases and related enzymes. Eur. J. Biochem. **200**:537-543.
17. **Reid, M. F. and C. A. Fewson.** 1994. Molecular characterization of microbial alcohol dehydrogenases. Crit. Rev. Microbiol. **20**:13-56.
18. **Saitou, R. and R. Nei.** 1987. A Neigbour-joining method: A new method for reconstructing phylogenetic trees. Mol. Biol. Evol. **44**:406-425.
19. **Sambrook, J., E. F. Fritsch and T. Maniatis.** 1989. Molecular Cloning. A Laboratory Manual, 2nd ed. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.
20. **Schönheit, P. and T. Schäfer.** 1995. Metabolism of hyperthermophiles. World J. Microbiol. Biotechn. **11**:26-57.
21. **van de Peer, Y. and R. de Wachter.** 1994. Treecon for Windows: A software package for teh construction and drawing of evolutionary trees for Microsoft Windows environment. Comput. Applic. Biosci. **10**:569-570.
22. **Voorhorst, W. G. B., R. I. L. Eggen, E. J. Luesink and W. M. de Vos.** 1995. Characterization of the *celB* gene coding for β-glucosidase from the hyperthermophilic archaeon *Pyrococcus furiosus* and its expression and site-directed mutation in *Escherichia coli*. J. Bacteriol. **177**:7105-7111.
23. **Voorhorst, W. G. B., Y. Guenguen, G. Schut, I. Dahlke, M. Thomm, J. van der Oost and W. M. de Vos.** 1997. Transcriptional regulation in the hyperthermophilic archaeon *Pyrococcus furiosus*: Coordinated expression of divergently transcribed genes in response to β-linked glucose polymers. Submitted for publication

Chapter 4

Cloning and Sequencing of the *lamA* Gene from the Hyperthermophilic Archaeon *Pyrococcus furiosus* Coding for an Endo-β-1,3-Glucanase and Its Expression and Site-directed Mutation in *Escherichia coli*

Yannick Gueguen, Wilfried G. B. Voorhorst, John van der Oost, and Willem M. de Vos

Journal of Biological Chemistry, 1997, 272, 31258-31264

Abstract

We report the molecular characterization of the first endo-β-1,3-glucanase from an archaeon. *Pyrococcus furiosus* is a hyperthermophilic archaeon that is capable of saccharolytic growth. The isolated *lamA* gene encodes an extracellular enzyme that shares homology with both endo-β-1,3- and endo-β-1,3-1,4-glucanases of the glycosyl hydrolase family 16. After deletion of the N-terminal leader sequence, a *lamA* fragment encoding an active endo-β-1,3-glucanase was overexpressed in *E. coli* using the T7-expression system. The *P. furiosus* endoglucanase was purified and showed highest activity on the β-1,3 glucose polymer laminarin, but has also activity on the β-1,3-1,4 glucose polymers lichenan and β-glucan. The enzyme shows optimal activity between 100 and 105 °C and was demonstrated to be the most thermostable endo-β-1,3-glucanase described up to now. Site-directed mutagenesis was performed to study the role of an additional methionine residue, in the predicted active site, found only in glycosyl hydrolases of the family 16 that show predominantly endo-β-1,3-glucanase activity. The deletion of this methionine did not change the substrate specificity of the endoglucanase, but it did cause a severe reduction in its catalytic activity, suggesting a structural role of this residue in constituting the active site. HPLC analysis showed *in vitro* hydrolysis of laminarin by the endo-β-1,3-glucanase proceeds more efficiently in combination with an exo-β-glycosidase from *Pyrococcus furiosus* (CelB). This most probably reflects the physiological role of both enzymes: co-operation during growth of *Pyrococcus furiosus* on β-glucans.

Introduction

β-1,3-Glucanases are widely distributed among bacteria, fungi and higher plants. They are classified as exo-β-1,3-glucanases (β-1,3-glucan glucanohydrolase [EC 3.2.1.58]) and endo-β-1,3-glucanases (β-1,3-glucan glucanohydrolase [EC 3.2.1.6 or EC 3.2.1.39]). Distinct physiological roles of β-1,3-glucanases have been proposed. In plants, involvement in cell differentiation and defence against fungal pathogens has been suggested (1). In fungi, β-1,3-glucanases seem to have different functions in morphogenetic processes, β-glucan mobilization, and fungal pathogen-plant interactions (2). Recently, the first metazoan β-1,3-glucanase, that may be involved in the early embryogenesis, has been described (3). In bacteria, a metabolic function has been reported for endo-β-1,3-glucanases and endo-β-1,3-1,4-glucanases (4). Both types of bacterial enzymes are polysaccharide endohydrolases with closely related

specificities (5,6). β-1,3-glucanases hydrolyse 1,3-β-glucosyl linkages, but usually require a region of unsubstituted, contiguous 1,3-β-linked glucosyl residues. In contrast, β-1,3-1,4-glucanases catalyse the hydrolysis of 1,4-β-glucosyl linkages, only when the glucosyl residue itself is linked at the (O)3 position (7).

Genes encoding bacterial β-1,3 and β-1,3-1,4-glucanases have been cloned and sequenced from different *Bacillus* species (8-16), *Fibrobacter succinogenes* (17, 18), *Cellvibrio mixtus* (19), *Thermotoga neapolitana* (20), *Ruminococcus flavefaciens* (21, 22), *Oerskovia xanthineolytica* (23), *Clostridium thermocellum* (24, 25, 26, 27), and *Rhodothermus marinus* (28). All bacterial endo-β-1,3-glucanases (laminarases) known to date share sequence similarity with endo-β-1,3-1,4-glucanases (lichenases) and have been classified in the same family 16 of glycosyl hydrolases (29, 30, 31). On the other hand, eukaryal endo-β-1,3-1,4-glucanases and endo-β-1,3-glucanases have been classified in the family 17 of glycosyl hydrolases. However, the first metazoan β-1,3-glucanase, described from a sea urchin, shares homology with both β-1,3 and β-1,3-1,4-glucanases of glycosyl hydrolases family 16 (3).

Presently, no β-specific endoglucanases have been reported in the Archaea, the third domain of life (32, 37). In this study, we report the characterization of an endo-β-1,3-glucanase from *Pyrococcus furiosus*, a heterotrophic hyperthermophilic archaeon that is able to grow optimally at 100°C on a wide range of carbohydrate substrates including starch, maltose, cellobiose (32, 33). *P. furiosus* was found to contain several hydrolytic enzyme activities related to sugars degradation (34) and utilizes a modified Embden-Meyerhof Pathway, which involves at least three unique enzymes: two ADP-dependent kinases (35), and a glyceraldehyde-3-phosphate: ferredoxin oxidoreductase (36). However, *P. furiosus* cannot grow on cellulose or carboxymethylcellulose and nothing has been described concerning the growth on other β-linked carbohydrates substrates such as laminarin, β-glucan and lichenan (32, 37).

The *P. furiosus lamA* gene was characterized and found to encode a secreted endo-β-1,3-glucanase belonging to the glycosyl hydrolyse family 16. Further characterization of the overproduced catalytic domain of this enzyme revealed that it was the most thermostable endo-β-1,3-glucanase detected so far. A mutant form of the catalytic domain was obtained after site directed mutagenesis of the *lamA* gene and revealed that a conserved methionine residue is required for full activity. Based on the specificity of the endo-β-1,3-glucanase and the here described capacity of *P. furiosus* to grow on laminarin, a role for this extracellular enzyme during the fermentation of β-1,3-linked glucose polymer is proposed.

Experimental procedures

Organisms and growth conditions- *P. furiosus* (DSM 3638) was used in this study (32). *E. coli* BL21 (DE3) harboring pLysE was used as host strain for the recombinant plasmid of pGEF (38, 39). *E. coli* TG1 was also used as host strain for the cloning vector pLUW530 and pLUW531 (40). *E. coli* was grown in L broth medium in a rotary shaker at 37°C. Ampicillin was added to L broth in a final concentration of 100 μg/ml. Isopropyl-β-D-thiogalactopyranoside (IPTG) was added in a final concentration of 0.4 mM, for the induction of gene expression.

Cloning of the endo-β-1,3-glucanase gene- A 4.8 kb fragment containing the *lamA* gene coding for *P. furiosus* endo-β-1,3-glucanase was cloned into pTZ19R (41). Based on its sequence, primers were designed to amplify the *lamA* gene by the polymerase chain reaction on a DNA Thermal Cycler (Perkin Elmer Cetus, Norwalk, Conn.) (Fig. 1). The two primers with *Nco*I and *BamH*I restriction sites (in bold),were the following: BG 194 (5'-GCG CG**C CAT GG**T CCC TGA AGT GAT AGA AAT AGAT, sense) and BG 195 (5'-CGC GC**G GAT CC**T CAA CCA CTA ACG AAT GAG TA, antisense). In addition to the template and the primers, the 100 μl reaction mixture contained 0.2 mM dNTPs, *Taq* DNA polymerase buffer, 5 mM $MgCl_2$, 5 U *Taq* DNA polymerase (Life Technologies Inc.) and was subjected to 30 cycles of amplification (30 s at 94°C, 30 s at 50°C and 30 s at 72°C). The putative signal peptide of the endoglucanase was excluded to produce the mature enzyme in the cytoplasm of *E. coli*. A PCR product with the expected size was digested with *Nco*I and *BamH*I and cloned in a pGEF vector (38), resulting in pLUW530, and transformed into *E. coli* TG1 and BL21 (DE3), pLysE using standard procedures (42).

Overexpression of the lamA gene and purification of the endo-β-1,3-glucanase- An overnight culture of *E. coli* BL21 (DE3), pLysE harboring pLUW530 was diluted 1:20 and grown until the A_{600} reached 0.6. The culture was induced with 0.4 mM of IPTG for 4 h. Cells were harvested by centrifugation, resuspended in 100mM sodium citrate buffer (pH 5.5) and sonicated using a Branson sonifier. Cell debris was removed by centrifugation (10,000 x g for 10 min). The resulting supernatant was heated for 30 min at 70°C and precipitated proteins were removed by another centrifugation. The supernatant was subsequently dialyzed against Tris-HCl buffer (20 mM, pH 8.0), loaded on a Q-Sepharose Fast Flow (Pharmacia) column (5 x

25 cm) that was equilibrated with the same buffer. Bound proteins were eluted by a linear gradient of NaCl (0 to 1 M in Tris-HCl buffer). Active fractions eluted around 0.5 M NaCl.

Protein samples were analysed by sodium dodecyl sulphate polyacrylamide gel electrophoresis (SDS-PAGE) using the method of Laemmli (43). Proteins samples for SDS-PAGE were prepared by heating for 10 min at 110°C in an equal volume of sample buffer (0.1 M citrate-phosphate buffer, 5% SDS, 0.9% 2-mercaptoethanol, 20% glycerol, pH 6.8).

β-Glucanase assay, protein determination and western blot- The standard assay for β-glucanase activity was carried out at 80°C, for 10 min, using 0.5% (w/v) laminarin (Sigma Chemical Co.) as substrate in 0.1 M phosphate (NaH_2PO_4/K_2HPO_4) buffer (pH 6.5). The reducing sugars released were detected by the dinitrosalicylic acid (DNS) method, with glucose as a standard (44). Enzyme and substrate blanks were also included. 1 U enzyme activity is defined as the amount of enzyme required to release 1 μmol of reducing sugars/min. Protein concentrations were determined by the method of Bradford, with BSA as the standard (45).

Antibodies were raised against LamA by injecting enzyme purified from *E. coli* into rabbits. The proteins were separated by SDS-PAGE and blotted onto nitrocellulose filters. These were incubated with serum containing antibodies (Dilution 1/1000) raised against LamA. The reaction of proteins with the antibodies was detected using the western blot Alkaline Phosphatase system (Promega).

Temperature and pH optima, thermal stability, kinetic parameters- The temperature optimum was determined by running the standard assay at temperatures from 40°C to 115°C. The pH optimum of the enzyme was determined by running the standard assay at 80°C, using citrate-phosphate (citric acid/NaH_2PO_4) buffer (0.1 M) and phosphate (NaH_2PO_4/K_2HPO_4) buffer (0.1M) for the pH range 4-7 and 6-8, respectively. Thermostability was determined using diluted enzyme (0.14 mg/ml pure enzyme was diluted 10 times in 200 mM Tris-HCl buffer, pH 6.5), incubated at 80°C, 90°C and 100°C. Samples were taken at time intervals and residual activity was determined by the standard assay at 80°C. Michaelis-Menten constants were determined from Lineweaver-Burk representations of data obtained by determining the initial rate of laminarin and lichenan hydrolysis under the assay conditions described above and using a range of 0.25 to 10 mg of substrate per ml.

Mutagenesis of the lamA gene- The plasmid pLUW500 carrying the *lamA* gene was used as a template for mutagenesis following the splicing by overlap extension PCR (46). The first step comprised the use of primers introducing the mutation and the two flanking primers (BG 194 and BG195), described above. The two mutagenesis primers were the following: BG 227 (5'-GCC AAG **GAA TTC** TAT GTC TAT TTC TCC ACA ATT TGG C, antisense) and BG 226 (5'-GAA ATA GAC ATA **GAA TTC** CTT GGC CAT GAG CAA, sense). The second step used the overlapping products of the first two PCR reactions as a template, and the flanking primers (BG 194, BG 195) to yield the *lamA* gene with the mutation. In order to simplify detection of mutants, an *Eco*RI site (in bold in the primers) was created. The amplified DNA was cloned as *Hin*dIII/*Bam*HI fragment, into the *lamA* gene in pLUW530. The mutation was checked by *Eco*RI digestion and DNA sequence analysis. The resulted plasmid (pLUW531) was transformed to *E. coli* BL21 (DE3), pLysE for overproduction of the protein.

DNA sequencing- Nucleotide sequencing was performed by the dideoxynucleotide chain termination method (47) with a Li-Cor automatic sequencing system (model 4000L). Computer analysis of sequencing data was performed by using the PC/GENE program version 5.01 (IntelliGenetics Inc., Moutain view, Calif.) and the GCG package version 7.0 at the CAOS/CAMM Centre of the University of Nijmegen.

Nucleotide sequence accession number- The nucleotide sequence reported has been submitted to the GenBank/EMBL Data Bank with the accession number AF013169.

Results

***Cloning and sequencing of the lamA gene encoding an endo-β-1,3-glucanase from* P. furiosus-** The nucleotide sequence of a part of the 4.8 kb DNA insert from pLUW500 (41, 48) revealed the presence of an 894 bp open reading frame (Fig. 1). The encoded 297-amino acid protein is homologous to a family bacterial β-1,3 and β-1,3-1,4-endoglucanases (see below) and contains a putative signal peptide of 20 residues (49). The molecular mass of the predicted mature protein is 31,550 Da

and the estimated isoelectric point 5.15. In order to overproduce this β-glucanase, the *lamA* gene was amplified by the PCR method and cloned into pGEF (38) under control of the T7 promotor (39). The resulting plasmid, containing the catalytic domain of the endo-β-1,3-glucanase (starting at residue 35 and called LamA), was named pLUW530. The plamid pLUW530 harboring *lamA* was checked by DNA sequence analysis.

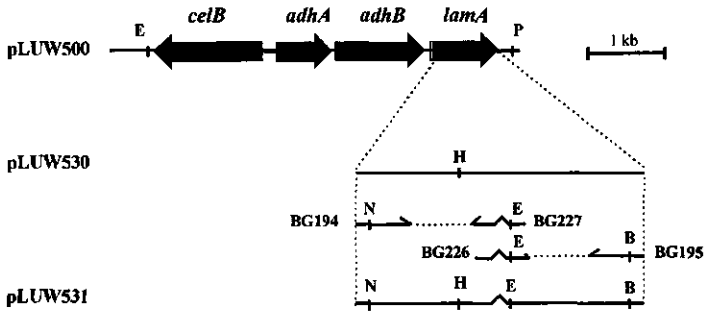


Figure 1. Genetic organisation of the *lamA* region. This DNA fragment was originally cloned as a fragment from chromosomal DNA of *P. furiosus* and inserted into the EcoRI / PstI sites of pTZ19R (41). Primers BG194 and BG195 used in the cloning of the *lamA* gene and primers BG226 and BG227 used in site-directed mutagenesis of LamA are indicated by arrows. Relevant restriction sites are indicated. H, *Hind*III; E, *Eco*RI; N, *Nco*I; B, *Bam*HI; P, *Pst*I.

Primary structure comparison- Comparison of the deduced primary strucutre of *lamA* with enzymes present in the GenBank database indicated highest similarity with representatives of the family 16 of glycosyl hydrolases, which includes bacterial endo-β-1,3- and endo-β-1,3-1,4-glucanases, *Streptomyces coelicolor* agarase and *Alteromonas carrageenovora* κ-carrageenase. Scores of 52.4 and 46.9 % identity were found with the β-glucanases of *T. neapolitana* and *O. xanthineolytica*, respectively. Multiple alignments of the deduced amino acid sequence of *P. furiosus lamA* and other β-glucanases showed many conserved amino acids (Fig. 2) including glutamate 170 and glutamate 175 (*P. furiosus* amino acid numbering is used throughout the paper). The results of structural analysis, mutagenesis (50, 51, 52, 53) combined with inhibitor attachment (52, 54) studies of *Bacillus* β-glucanases showed the direct involvement of these two carboxylic acids in catalysis. Glutamate 175 corresponds to the general acid and glutamate 170 to the catalytic nucleophile or plays a role in

electrostatic stabilization of an oxocarbonium intermediate ion (55). Moreover, the crystal structure of the hybrid *Bacillus* β-1,3-1,4-β-glucanase (52) showed the presence of a calcium binding site on the surface of the enzyme. Experiments demonstrated that calcium, bound to the hybrid *Bacillus* enzyme, stabilizes the three-dimensional structure of the protein leading to increased thermal stability (56, 57). The calcium atom is bound (52) on the convex face of the protein, to the backbone carbonyl oxygens atoms of proline *[9]*, glycine 78 *[45]* and aspartate 287 *[207]* and to the carboxylate oxygen of aspartate 287 *[207]* (italic numbers correspond to *Bacillus* hybrid β-glucanase numbering (52)). This aspartate 287 residue is conserved among most of the β-glucanases of the family 16 (Fig. 2). The *P. furiosus* β-1,3-glucanase also does contain this aspartate residue, and indeed was found to be protected by calcium against thermal inactivation (data not shown).

Figure. 2. Alignment of *P. furiosus* LamA with other members of family 16 of glycosyl hydrolases. Sequences were deduced from the following accession numbers: *T. neapolitana* laminarase (Z47974); *O. xanthineolytica* β-1,3-glucanase (U56935); *C. thermocellum* 1 β-1,3(4)-glucanase, licA (X89732); *R. marinus* β-glucanase (U04836); *B. licheniformis* β-1,3-1,4-glucanase (X57279); *B. subtilis* β-glucanase (X00754); *B. macerans* β-1,3-1,4-glucanase (Z25874); *C. thermocellum* 2 β-1,3-1,4-glucanase, licB (X63355); *B. brevis* β-1,3-1,4-glucanase (M84339). The numbering of the *P. furiosus* LamA is indicated. Conserved residues are indicated in bold. The asterisks above the sequences indicate the conserved glutamate residues that have been identified as acid-base catalyst and active site nucleophile. Underlined and boxed regions correspond to β-strands and α-helix, respectively, according to *B. licheniformis/B. macerans* hybrid β-1,3-1,4-glucanase secondary structure (50, 52).

Overexpression of the* lamA *gene in E. coli and purification of the endo-β-1,3-glucanase- In order to characterize *P furiosus* endo-β-1,3-glucanase, we overexpressed its catalytic domain, hereafter called LamA, in *E coli* BL21 (DE3), pLysE. The *E. coli* strain harboring the plasmid pLUW530 was grown until mid-log phase and induced by IPTG. SDS-PAGE analysis of cell free extract of induced cells revealed an additional band of 30 kDa corresponding with the calculated molecular mass of the *lamA* product (Fig. 3). The 30 kDa band was absent in extracts of *E. coli* BL21 (DE3), pLysE carrying the pGEF plasmid without insert (Fig. 3). The level of LamA production in *E. coli* harboring pLUW530 is 150 U/mg of total protein. Using this expression system, LamA could be produced up to 15% of the total cell protein. LamA was soluble and no indication of the formation of inclusion bodies was found. LamA could easily be purified to homogeneity by a two-step purification procedure consisting of a heat incubation (30 min at 70°C) that results in the denaturation of the majority of the *E. coli* proteins, followed by an anion exchange chromatography step (data not shown). About 7-fold purification was obtained and the isolated enzyme was estimated to be at least 95% pure by SDS-PAGE (Fig. 3).

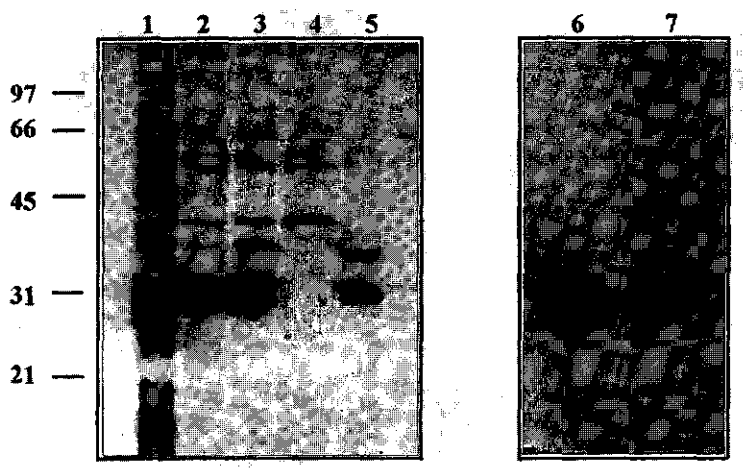


Figure. 3. A- SDS-PAGE of *E. coli* BL21 (DE3), pLysE containing the plasmid pLUW530, overproducing the catalytic domain of *P. furiosus* LamA. lane 1, crude cell extract; lane 2, soluble fraction; lane 3, supernatant after heat incubation (30 min at 70°C); lane 5, purified fraction after anion exchange chromatography. lane 4, *E.coli* harboring pGEF supernatant after heat incubation.
B- Western blot analysis. lane 6, *E. coli* crude extract; lane 7, *P. furiosus* supernatant culture grown on laminarin. Proteins were probed with anti-LamA antibodies.

Endo-β-1,3-glucanase characterization- The specific enzyme activities of the purified LamA β-glucanase with various substrates are summarized in Table 1. Control extracts from *E. coli* carrying pGEF did not degrade the tested substrates. These results indicate that LamA has highest specificity for the soluble β-1,3-glucan laminarin. A 10-fold lower activity was observed on β-1,3-1,4-linked glucans, like lichenan and barley β-glucan. The β-1,4-linked substrates (carboxymethyl cellulose, cellulose and xylan) were not hydrolysed. Based on the substrate specificity, *P. furiosus* LamA enzyme produced in *E. coli* was characterized as β-1,3-glucanase (laminarase, EC 3.2.1.39). The mode of action of LamA was examined by analyzing the corresponding hydrolysis products by HPLC (Fig 4.). The enzyme splitted laminarin randomly, yielding glucose, laminaribiose and higher laminari-oligosaccharides. These results confirm that LamA is an endo-β-1,3-glucanase. The combined action of LamA and CelB (an exo-β-glycosidase from *P. furiosus,* (32, 41)) resulted in the almost complete hydrolysis of laminarin (96%) to glucose. The incubation of laminarin in the same conditions with CelB only, resulted in the degradation of 5% of the laminarin, with only glucose as end product.

Table I

Substrate specificity of the purified endo-β-1,3-glucanase of P. furiosus

Enzyme activity was measured in 100 mM phosphate buffer, at pH 6.5 for 5 min. All substrates were used at a concentration of 5 mg/ml. Each assay was performed with 0.68 μg of protein per ml. The amount of reducing sugars released were detected by the dinitrosalicylic acid (DNS) method (44). Activity using PNPC (*p*-nitrophenyl-β-D-cellobioside) was determined by using the release of *p*NP by adsorbance at 405 nm.

Substrate	Main linkage type (monomer)[a]	Specific activity (U/mg)
laminarin	β-1,3(Glc)	922
lichenan	β-1,3;1,4(Glc)	95
barley ß-glucan	β-1,3;1,4(Glc)	99
PNPC	β-1,4(Glc)	n.d.[b]
cellulose	β-1,4(Glc)	n.d.
CM cellulose	β-1,4(Glc)	n.d.
xylan	β-1,4(Xyl)	n.d.
chitin	β-1,4(GlcNAc)	n.d.

[a] Glc, glucose; GlcNAc, glucose acetylamine; Xyl, xylose. [b] n.d. : non detectable

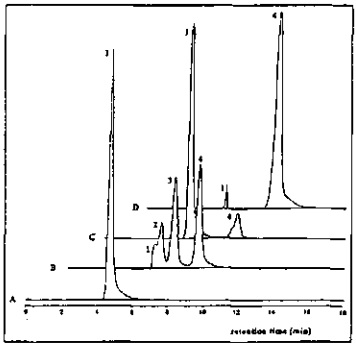


Figure 4. HPLC analysis of *P. furiosus* LamA and CelB action on laminarin (0.5% w/v). (A) laminarin control, incubation at 80°C for 4 hours; (B) incubation with LamA (5 μg/ml), at 80°C for 4 hours; (C) incubation with CelB (5 μg/ml), at 80°C for 4 hours; (D) incubation with LamA (5 μg/ml) and CelB(5 μg/ml), at 80°C for 4 hours. Samples were analysed on an Aminex HPX-87-H column (Biorad). Identified products are: **1**-laminarin; **2**-laminari-oligomers; **3**-laminaribiose; **4**-glucose.

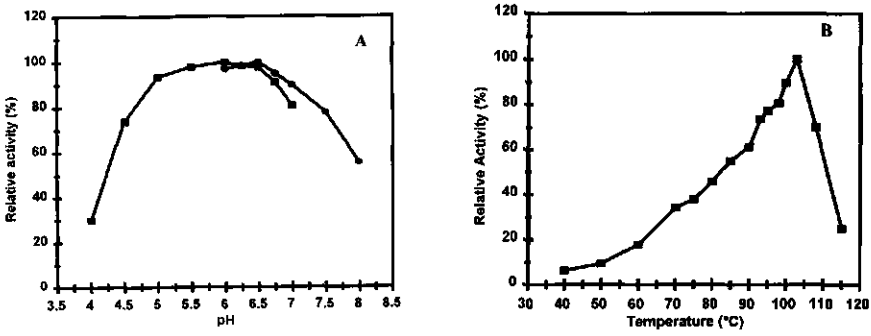


Figure 5. Influence of pH and temperature on the activity of the purified endo-β-1,3-glucanase of *P. furiosus*. (A) Enzyme activity was measured, at 80°C for 5 min using laminarin as substrate, in 100 mM citrate-phosphate buffer (■) at pH 4 to 7 and in 100 mM phosphate buffer (●) at pH 6 to 8. The 100% activity corresponds to 935 U/mg of protein. (B) Enzyme activity was measured in 100 mM phosphate buffer, at pH 6.5 for 5 min, using laminarin as substrate. The 100% activity corresponds to 1842 U/mg of protein. Each assays were performed with 0.68 μg of protein per ml.

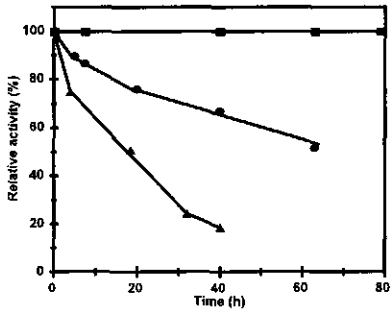


Figure 6. Thermal stability of the *P. furiosus* endo-β-1,3-glucanase. The enzyme was pre-incubated at 80°C (■), 90°C (●) and 100°C (▲), in 200 mM Tris-HCl buffer, pH 6.5. Residual activity was measured in 100 mM phosphate buffer, at pH 6.5 for 5 min, using laminarin as substrate. The 100% activity corresponds to 945 U/mg of protein.

The laminarase enzyme activity was investigated at different pH values and temperatures (Fig. 5). The pH optimum was found to be 6-6.5. The enzyme exhibited at least 80% of its optimal activity over a rather broad pH range, from 5 to 7 (Fig. 5A). The temperature for maximum activity was 100-105°C. The enzyme was almost inactive at 40°C and began to show significant activity above 60°C (Fig. 5B). Thermal stability was investigated by incubating the pure enzyme up to 110 h at different temperatures (Fig. 6). The enzyme retained 100% of its laminarase activity after 110 h incubation at 80°C. At higher temperatures, thermostability decreased: at 90°C and 100°C, the enzyme showed a half life of 64 h and 19 h, respectively. At 110°C, the enzyme was rapidly inactivated, with a half life of only 15 min (data not shown).

Site-directed mutagenesis of the* P. furiosus *endo-β-1,3-glucanase- To investigate whether the deletion of methionine at position 174 in LamA, would affect the β-1,3 substrate specificity of the enzyme, a mutated *lamA* gene was constructed in plasmid pLUW530. Relative activities and kinetic characteristics of mutant and wild-type enzymes were determined with laminarin and lichenan as substrates (Table 2). The specific activities of the mutant LamA, on both laminarin and lichenan, are about 10-fold lower than respective wild-type values. Catalytic parameters were determined at 80°C and pH 6.5 (optimal conditions for the wild-type enzyme). The methionine deletion led to an enzyme with 13.1% of wild-type V_{max} and a slightly higher K_m , on laminarin as substrate. With lichenan as substrate, the mutant LamA has 7.9% of the wild-type V_{max} and the same K_m value.

***Detection of the endo-β-1,3-glucanase in* P. furiosus-** The unexpected discovery of the endo-β-1,3-glucanase in *P. furiosus* prompted us to study the capacity of this hyperthermophilic archaeon to grow on β-1,3-linked glucose polymers. Remarkably, *P. furiosus* was found to grow on laminarin (0.3% w/v) and lichenan (0.3% w/v) as substrate (data not shown) and endo-β-1,3-glucanase activity was detected in the culture medium. Western blot analysis with immune serum directed against LamA gave an immunoreactive band with the culture supernatant of *P. furiosus,* grown on laminarin as substrate (Fig. 3). The size of the product from *P. furiosus* is slightly higher (around 1 kDa) than the one produced in *E. coli*, which is in agreement with the fact that, we have deleted, during the overexpression step, a N-terminal fragment of 34 amino acid which is larger than the predicted signal peptide of 20 amino acid.

Table II

Enzyme activity and kinetic parameters of wild type and mutant LamA of P. furiosus.

Lineweaver-Burk plots were used to determine the kinetic constants. Activity on laminarin and lichenan were measured in 100 mM phosphate buffer, at pH 6.5 for 5 min. Substrates were used at different initial concentrations. Each assay was performed with 0.68 µg of protein per ml. The amount of reducing sugars released were detected by the dinitrosalicylic acid (DNS) method (44).

Substrate	Enzyme	Specific Activity [a] (U/mg)	Activity [b] (%)	K_m (mg/ml)	V_{max} [b] (%)
laminarin					
β-1,3(Glc)	wt	1073	100	2.8	100
	mutant	111.6	10.4	3.5	13.1
lichenan					
β-1,3-1,4(Glc)	wt	85	100	4.7	100
	mutant	7.1	8.3	4.9	7.9

[a] Enzyme activity was measured in 100 mM phosphate buffer pH 6.5, at 80°C, for 10 min.
[b] Percentages of residual activity and V_{max} are expressed relative to the wild type for each substrate.

Discussion

In this paper, we describe the overexpression of the *lamA* gene which encodes a extremely thermostable endo-β-1,3-glucanase from the hyperthermophilic archaeon *P. furiosus.* The overproduction and purification of the endo-β-1,3-glucanase LamA was carried out by cloning a *lamA* gene fragment into the T7 expression system, after deletion of the hydrophobic N-terminal sequence (34 amino acids). This is the first description of an endo-β-1,3-glucanase from an archaeon and it is the most thermostable endo-β-1,3-glucanase known up to now. The second most thermostable endoglucanase of the family 16 of glycosyl hydrolases characterized, is that of the thermophilic bacterium *R. marinus,* which has an optimum temperature of 85°C and has 85% residual activity after 16 h incubation at 80°C. The optimum temperature of the *P. furiosus* β-glucanase is 100-105°C and the enzyme retained 100% and 50% of its activity after incubation for 85 h at 80°C and 16 h at 100°C, respectively. The optimum temperature of the enzyme is near the optimal growth temperature (100°C) of the *P. furiosus* strain (35). The enzyme activity has a broad pH optimum around pH 6, which is comparable to other endo-β-1,3-glucanases and endo-β-1,3-1,4-

glucanases, with the exception of the *B. brevis* enzyme, which has optimal activity at pH 9.0. The *P. furiosus* β-glucanase was able to hydrolyse both β-1,3-glucan and β-1,3-1,4-glucan, but not the β-1,4-linked substrates like cellulose. Bacterial laminarases and lichenases often yield trisaccharides and tetrasaccharides as the main degradation products (58). On the other hand, as other bacterial and fungal endo-β-1,3-glucanases (2, 26, 28), the endo-β-1,3-glucanase of *P. furiosus*, hydrolyses β-1,3-glucans with glucose, laminaribiose, and laminaritriose as end products.

All the bacterial laminarases and lichenases sequenced so far have been classified in the same family 16 of glycosyl hydrolases. The mechanism for β-glucanase action must take into account two general considerations (55, 59). The hydrolysis of the glycosyl bond can proceed via either retention or inversion of the configuration at the anomeric carbon atom. Since NMR analysis has indicated retention of configuration in case of the *B. licheniformis* β-glucanase (60), the same stereochemical course is assumed for all other members of the family (61). This mechanism requires two functional groups of the enzyme to be present in the appropriate spatial setting. The first either acting as a nucleophile or by providing electrostatic stabilization, and the second as a general acid. Several experiments with β-glucanases from different *Bacilli* (54, 55) demonstated the role of glutamate 170 as the catalytic residue involved in the nucleophilic attack on the substrate. Glutamate 175 was demonstrated to be the most likely candidate to function as the general catalyst (50, 51). As demonstrated by the crystal structure of the *B. macerans* enzyme and a substrate analogue (50, 52, 62), the β-glucanase substrate binds to a pronounced channel on the molecular surface, in which four residues, glutamate 170, aspartate 172, glutamate 175, and glycine 178, are located on the same β-strand (52). These residues are completely conserved in all available sequences of β-glucanases (Fig. 2). Interestingly, the alignment suggests that amino acid sequences surrounding the active site can be divided into two categories (Fig. 2). From glutamate 170 to phenylalanine 176, there is a strict alternation of polar and nonpolar side chains in the catalytic sites of the endo-β-1,3-1,4-glucanases. This organization is different in the endo-β-1,3-glucanases because of the insertion of an extra methionine residue between isoleucine 173 and glutamate 175. Assuming that both the active site structure and the catalytic mechanism are similar in a same family, this residue has been proposed to form a β-buldge, introducing a kink in the β-strand and thereby allowing the methionine side chain to point towards the hydrophobic core and the glutamate 175 to participate in catalysis (50, 63). This structural rearrangement of the active site most likely will

affect catalysis. Morever, substrate specificity analysis within the glycosyl hydrolyse family 16 shows that these enzymes can be classified in two subfamilies : (i) a group of β-glucanases able to hydrolyse β-1,3-glucan or both β-1,3 and β-1,3-1-4-glucan (such in the case of *P. furiosus* and *R. marinus*) and (ii) a group of β-1,3-1-4-glucanases. The structural change brought about by the inserted methionine has been correlated with the altered specificity, from β-1,3-1,4 linked substrates to β-1,3 linkages ones (28, 50). The role of this extra methionine in *P. furiosus* β-glucanase was investigated by site-directed mutagenesis. Removal of methionine 174 in the *P. furiosus* β-glucanase resulted in a decreased enzyme activity by a factor of 10 for both laminarin and lichenan. However, we did not observe a shift of relative affinity of the mutant enzyme from β-1,3-glucan towards β-1,3-1,4-glucan substrates. Hence, deletion of methionine 174 affected mostly V_{max} , but not K_m. The results suggest that the methionine 174 residue may assist in catalysis as a neighbouring group to the essential glutamate 175 and suggest a structural role of this residue in the constitution of the active site. However, it does not significantly contribute to specific substrate binding, as seen by the unchanged K_m upon deletion. Moreover, studies of a mutant *B. macerans* β-glucanase in which a methionine has been inserted in the active site region show it to be enzymatically inactive (50). These results suggest that further modifications are required to extend the substrate specificity of β-1,3-glucanases, for example amino acids responsible for the binding of the different polyaccharide substrates. Ferrer and co-workers (23) noted that conserved regions of the β-glucanases family are rich in aromatic residues, tryptophan in particular. This residue was demonstrated to play a role in substrate binding of some cellulases (64). Furthermore, a limited number of tryptophan residues is invariant either in β-1,3-glucanases or in β-1,3-1,4-glucanases [in positions 46, 50, 89, 150, 154, 257, 270], suggesting a possible function in β-glucanase substrate specificity of these residues. Moreover, Juncosa and co-workers (51) proposed, after site directed mutagenesis on *Bacillus licheniformis* β-1,3-1,4-glucanase, a possible role in substrate binding of two glutamate residues (in position 122 and 203), which are conserved only among β-1,3-1,4-glucanases. A better understanding of the basis of the β-1,3- and β-1,3-1,4 β-glucanase substrate specificity awaits a detailed comparison of the three dimensional structures of these two classes of enzymes and experimental verifications of the derived conclusions by protein engineering.

Previously, it has been described that *P. furiosus* is capable to degrade α-linked glucose polymers like starch and glycogen by the concerted action of α-

amylase, pullulanase and α-glucosidase (65). Here, we have demonstrated that, in addition, *P. furiosus* is able to grow on β-linked glucose polymers. Two enzymes are involved in the hydrolysis of laminarin to glucose : the LamA encoded extracellular endo-β-1,3-glucanase and the β-glucosidase (CelB). The physiological substrate of the latter endo-β-1,3-glucanase may be a β-1,3-linked carbohydrate component of the cell wall of a variety of marine organisms, such as eukaryotic algae (laminarin) or methanogenic archaea (pseudopeptidoglycan) (66). As it has been demonstrated *in vitro*, the LamA and CelB enzymes are anticipated to catalyse the hydrolysis of sugars polymers to glucose in a cooperative manner. The extracellular endo-β-1,3-glucanase with high-level thermostability could function around the living area of *P. furiosus* degrading the β-1,3-glucose polymer to smaller oligomers, that are imported into the cell and hydrolysed to glucose by the exo-glycosidase (CelB). Thus, these two enzymes could start the catabolic pathway of *P. furiosus* growing on β-1,3-glucose polymer, that could ressemble the natural substrate of the strain in *in vivo* conditions. Morever, the *lamA* gene is located near the β-glucosidase *celB* gene and two alcohol dehydrogenases (Fig. 1) (48). Clustering of the *lamA* and *celB* genes may be of functional significance. The regulation of the *celB* and *lamA* genes in *P. furiosus* is currently under investigation (48).

Acknowledgements

The authors gratefully acknowledge Dr. J.P. Schanstra and Dr. D.B. Janssen (Groningen) for providing the expression vector pGEF.

References

1. Grenier, J., Potvin, C., and Asselin, A. (1993). *Plant Physiol.* **103**, 1277-1283
2. Cruz, de la J., Pintor-Toro, J. A., Benitez, T., Llobell, A., and Romero, L. C. (1995) *J. Bacteriol.* 177, 6937-6945
3. Batchman, E. S., and McClay, D.R. (1996) *Proc. Natl. Acad. Sci. USA.* **93**, 6808-6813
4. Watanabe, T., Kasahara, N., Aida, K., and Tanaka, H. (1992) *J. Bacteriol.* **174**, 186-190
5. Chen, L., Fincher, G. B., and Hoj, P.B. (1993) *J. Biol. Chem.* **268**, 13318-13326
6. Varghese, J. N., Garrett, T. P., Colman, P. M., Chen, L., Hoj, P. B., and Fincher, G. B. (1994) *Proc. Natl. Acad. Sci. U. S. A.* **91**, 2785-2789
7. Parrish, F. W., Perlin, A. S., and Reese, E.T. (1960) *Can. J. Chem.* **38**, 2094-2104
8. Borriss, R., Buettner, K., and Maentsaelae, P. (1990) *Mol. Gen. Genet.* **222**, 278-283
9. Bueno, A., Vazquez de Aldana, C. R., Correa, J., Villa, T. G., and de Rey, F. (1990) *J. Bacteriol.* **172**, 2160-2167

10. Fiske, M. J., Tobey-Fincher, K. L., and Fuchs, R. L. (1990) *J. Gen. Microbiol.* **136**, 2377-2383
11. Gosalbes, M. J., Perez-Gonzalez, J. A., Gonzalez, R., and Navarro, A. (1991) *J. Bacteriol.* **173**, 7705-7710
12. Hofemeister, J., Kurtz, A., Borriss, R., and Knowles, J. (1986) Gene **49**, 177-187
13. Lloberas, J., Perez-Pons, J. A., and Querol, E. (1991) *Eur. J. Biochem.* **197**, 337-343
14. Louw, M. E., Reid, S. J., and Watson, T. G. (1993) *Appl. Microbiol. Biotechnol.* **38**, 507-513
15. Murphy, N., McConnell, D. J., and Cantwell, B. A. (1984) *Nucleic Acids Res.* **12**, 5355-5367
16. Yahata, N., Watanabe, T., Nakamura, Y., Yamamoto, Y., Kamimiya, S., and Tanaka, H. (1990) *Gene* **86**, 113-117
17. Erfle, J. D., Teather, R. M., Wood, P. J., and Irvin, J. E. (1988) *Biochem. J.* **255**, 833-841
18. Irvin, J. E., and Teather, R. M. (1988) *Appl. Environ. Microbiol.* **54**, 2672-2676
19. Sakellaris, H., Pemberton, J. M., and Manners, J. M. (1993) *FEMS Microbiol. Lett.* **109**, 269-272
20. Dakhova, O. N., Kurepina, N. E., Zverlov, V. V., Svetlichnyi, V. A., and Velikodvorskaya, G. A. (1993) *Biochem. Biophys. Res. Commun.* **194**, 1359-1364
21. Flint, H. J., McPherson, C. A., and Bisset, J. (1989). *Appl. Environ. Microbiol.* **55**, 1230-1233
22. Flint, H. J., Martin, J., McPherson, C. A., Daniel, A. S., and Zhang, J. X. (1993). *J. Bacteriol.* **175**, 2943-2951
23. Ferrer, P., Halkier, T., Hedegaard, L., Savva, D., Diers, I., and Asenjo, J. A. (1996) *J. Bacteriol.* **178**, 4751-4757
24. Schwartz, W. H., Schimming, S., and Staudenbauer, W. L. (1988) *Biotechnol. Lett.* **101**, 225-230
25. Schimming, S., Schwarz, W. H., and Staudenbauer, W. L. (1991) *Biochem. Biophys. Res. Commun.***177**, 447-452
26. Schimming, S., Schwarz, W. H., and Staudenbauer, W. L. (1992) *Eur. J. Biochem.* **204**, 13-19
27. Zverlov, V. V., Laptev, D. A., Tishkov, V. I., and Velikodvorskaja, G. A. (1991) *Biochem Biophys. Res. Commun.* **181**, 507-512
28. Spilliaert, R., Hreggvidsson, G. O., Kristjansson, J. K., Eggertsson, G., and Palsdottir, A. (1994) *Eur. J. Biochem.* **224**, 923-930
29. Henrissat, B. (1991) *Biochem. J.* **280**, 309-316
30. Henrissat, B., and Bairoch, A. (1993) *Biochem. J.* **293**, 781-788
31. Henrissat, B., and Bairoch, A. (1996) *Biochem. J.* **316**, 695-696
32. Kengen, S. W., Luesink, E. J., Stams, A. J., and Zehnder, A. J. (1993) *Eur. J. Biochem.* **213**, 305-312
33. Fiala, G., and Stetter K. O. (1986) *Arch. Microbiol.* **145**, 56-61
34. Kengen, S. W. M., de Boek, F. A. M., van Loo, N. D., Dijkema, C., Stams, A. J. M., and de Vos, W. M. (1994) *J. Biol. Chem.* **269**, 17537-17541
35. Kengen, S. W. M., Stams, A. J. M., and de Vos, W. M. (1996) *FEMS Microbiol. Rev.* **18**, 119-137
36. Ma, K., Zhou, Z. H., and Adams, M. W. W. (1994) *FEMS Microbiol. Lett.* **122**, 245-250
37. Bauer, M. W., Bylina, E. J., Swanson, R. V., and Kelly, R. M. (1996) *J. Biol. Chem.* **271**, 23749-23755
38. Schanstra, J. P., Rink, R., Pries, F., and Janssen, D. B. (1993) *Protein Expression Purif.* **4**, 479-489
39. Studier, F. W., Rosenberg, A. H., Dunn, J. J., and Dubendorff, J. W. (1990) *Methods Enzymol.* **185**, 60-89
40. Gibson, T. J. (1984) Studies on the Epstein-Barr-Virus genome. Ph.D. thesis. Cambridge University, Cambridge
41. Voorhorst, W. G. B., Eggen, R. I. L., Luesink, E. J., and de Vos, W. M. (1995) *J. Bacteriol.* **177**,7105-7111
42. Sambrook, J., Fritsch, E. F., and Maniatis, T. (1989) *Molecular cloning*: A laboratory manual, 2nd Ed., Cold Spring Harbor Laboratory, Cold Spring Harbor, NY
43. Laemmli, U.K. (1970) *Nature.* **227**, 680-685
44. Summer, J. B., and Somers, G. F. (1949) Dinitrosalisylic method for glucose, Laboratory experiments in biological chemistry, Academic Press, NY
45. Bradford, M. M. (1976) *Anal Biochem.* **72**, 248-254

46. Ho, N. S., Pullen, J. K., Horton, R. M., Hunt H. D., and Pease,, L. R.. (1990) *DNA Protein Eng. Tech.* **2**, 50-55
47. Sanger, F., Nicklen, S., and Coulson, A. R. (1977). *Proc. Natl. Acad. Sci. USA* **74**, 5463-5467
48. Voorhorst, W. G. B. (1997). Article in preparation.
49. Gierasch, L. M. (1989) *Biochemistry* **28**, 923-930
50. Hahn, M., Olsen, O., Politz, O., Borriss, R., and Heinemann, U. (1995). *J. Biol. Chem.* **270**, 3081-3088
51. Juncosa, M., Pons, J., Dot, T., Querol, E., and Planas, A. (1994) *J. Biol. Chem.* **269**, 14530-14535
52. Keitel, T., Simon, O., Borriss, R., and Heinemann, U. (1993) *Proc. Natl. Acad. Sci. U. S. A.* **90**, 5287-5291
53. Planas, A., Juncosa, M., Lloberas, J., and Querol, E. (1992) *FEBS Lett.* **308**, 141-145
54. Hoj, P. B., Condron, R., Traeger, J. C., McAuliffe, J. C., and Stone, B. A. (1992) *J. Biol. Chem.* **267**, 25059-25066
55. Sinnot, M. L. (1990) *Chem. Rev.* **90**, 1171-1202
56. Keitel, T., Meldegaard, M., and Heinemann, U. (1994) *Eur. J. Biochem.* **222**, 203-214
57. Welfle, K., Misselwitz, R., Welfle, H., Politz, O., and Borriss, R. (1995) *Eur. J. Biochem.* **229**, 726-735
58. Bacon, J. C., Gordon, A. H., Jones, D., Taylor, I. F., and Webley, D. M. (1970) *Biochem. J.* **120**, 67-71
59. Svensson, B., and Soggard, M. (1993) *J. Biotechnol.* **29**, 1-37
60. Malet, C., Gimenez-barbero, J., Bernabe, M., Brosa, C., and Planas, A. (1993) *Biochem. J.* **296**, 753-758
61. Gebler, J., Gilkes, N. R., Claeyssens, M., Wilson, D. B., Beguin, P., Wakarchuk, W. W., Kilburn, D. G., Miller Jr, R. C., Warren, R. A., and Withers, S. G. (1992). *J. Biol. Chem.* **267**, 12559-12561
62. Hahn, M., Keitel, T., and Heinemann, U. (1995) *Eur. J. Biochem.* **232**, 849-858
63. Richardson, J. S., Getzoff, E. D., and Richardson, D. C. (1978) *Proc. Natl. Acad. Sci. USA.* **75**, 2574-2578
64. Din, N., Forsythe, I. J., Burtnick, L. D., Gilkes, N. R., Miller, R. C., Warren, R. A. J., and Kilburn, D. G. (1994) *Mol. Microbiol.* **11**, 747-755
65. Sunna, A., Moracci, M., and Antranikian, G. (1997) *Extremophiles* **1**, 2-13
66. König, H., Kandler, M., Jensen, M. and Rietschel, E. T. (1983) Hoppe-Seyler's Z. Physiol. Chem. **364**, 627-636

Chapter 5

Transcriptional Regulation in the Hyperthermophilic Archaeon *Pyrococcus furiosus*: Coordinated Expression of Divergently Transcribed Genes in Response to ß-Linked Glucose Polymers

Wilfried G.B. Voorhorst, Yannick Gueguen, Gerti Schut, Isabell Dahlke, Michael Thomm, John van der Oost, and Willem M. de Vos

submitted for publication

Abstract

The genetic organization, expression, and regulation of the *celB* locus of the hyperthermophilic archaeon *Pyrococcus furiosus* was analyzed. This locus includes the *celB* gene, coding for a ß-glucosidase, and a divergently orientated gene cluster with the order of *adhA-adhB-lamA* that encodes two alcohol dehydrogenases and an extracellular ß-1,3-endoglucanase. Northern blot analysis showed that this gene cluster forms an 2.8 kb operon, designated the *lamA* operon. The amounts of the ß-glucosidase and the ß-1,3-endoglucanase were found to be largely dependent on the carbon source present and were highest when *P.furiosus* was grown on the ß-linked glucose polymers cellobiose or laminarin. Northern blot analysis showed that the expression of the monocistronic *celB* gene was controlled at the transcriptional level and that induction occured within ten minutes after addition of cellobiose. The transcription of the *lamA* operon was found to be co-regulated with that of the divergent *celB* gene and was also induced by growth on ß-linked glucose polymers. The transcription initiation sites of the *celB* gene and the *lamA* operon were identified and their promoters showed a high degree of homology, including identical TATA box sequences. The same transcription initiation sites were also found using a *P.furiosus* cell-free transcription system. However, the relatively low *in vitro* transcription efficiency suggests the contribution of a positive regulator in *P.furiosus*. A putative regulator binding motif was identified in the intergenic region of *celB* and *lamA* that showed a striking homology to a bacterial FNR box.

Introduction

One of the most striking differences that underlies the phylogenetic divergence of *Archaea, Bacteria* and *Eucarya* is the process of gene expression. Whilst the genome size and gene organization into operons resembles the bacterial situation (Zillig *et al.*, 1993; Baumann *et al.*, 1995; Ciaramella *et al.*, 1995), the archaeal transcription initiation process shows a great deal of similarity with that of *Eucarya* (Baumann *et al.*, 1995; Langer *et al.*, 1995; Hausner *et al.*, 1996; Thomm, 1996). The archaeal RNA polymerase is structurally related to the RNA polymerases II and III of *Eucarya* (Langer *et al.*, 1995). The archaeal promoter, closely resembles the eucaryal RNA polymerase II promoters, with the archaeal TATA box equivalent to the eucaryal TATA-box located approximately 25 nucleotides upstream of the transcription initiation site (Reiter *et al.*,

1990; Hausner *et al.*, 1991; Palmer and Daniels, 1995). Various archaeal cell-free transcription systems have been established, which efficiently initiate transcription at sites that are also used *in vivo* (Frey *et al.*, 1990; Hüdepohl *et al.*, 1990; Hethke *et al.*, 1996), and require at least two transcription factors, the TATA-binding protein (TBP) and the transcription factor IIB (TFIIB), components homologous to the eucaryal counterparts (Hausner *et al.*, 1996; Thomm, 1996). Moreover, eucaryal TBPs were found to be functionally exchangeable with their archaeal homologs in archaeal *in vitro* transcription systems (Hethke and Thomm, 1997; Wettach *et al.*, 1995).

In contrast to the growing understanding of the transcription initiation machinery in *Archaea*, little is known about the molecular mechanisms controlling regulation of archaeal transcription. The few studies that address regulation of transcription have mainly focussed on describing modulations of gene expression in Crenarchaeotes and two subgroups of the Euryarchaeota, the halophiles and methanogens and their viruses (Zillig *et al.*, 1993; Zillig *et al.*, 1988; Reiter, *et al.*, 1988). Recently, it was found in thermophilic methanogens that the expression of genes involved in methane production is controlled at the transcriptional level in response to growth-phase and substrate levels (Morgan *et al.*, 1997; Nölling and Reeve, 1997).

Presently, no transcription regulators have been identified in *Archaea* (Zillig *et al.*, 1993; Ciaramella *et al.*, 1995). Gene specific regulators that interact with TBP and/or TFIIB, as in *Eucarya*, have been suggested for *Archaea* based on the regions of homology among their TBPs (Baumann *et al.*, 1995). However, recent structural analysis of the TBP of *Pyrococcus woesei* has revealed that the site for regulatory interaction differs from its eucaryal homolog, whereas the DNA-binding site is highly conserved (Dedecker *et al.*, 1996). It is therefore possible that bacterial regulators are functional in *Archaea*, which is supported by the identification of a number of homologs of bacterial regulators at the sequence level (Bult *et al.*, 1996; Fitz-Gibbon *et al.*, 1997). Moreover, two palindromic repressor binding sites have found to be functional in halophilic and methanogenic archaea, although the repressors have not yet been identified (Ken and Hackett, 1991; Stolt and Zillig, 1992; Cohen-Kupiec *et al.*, 1997).

In the present study, we extend the molecular analysis of archaeal transcription regulation with the third subgroup of the Euryarchaeota, the hyperthermophiles. The most extensively studied representative of this group is *Pyrococcus furiosus* (Fiala and Stetter, 1986), growing optimally at 100°C on a wide range of substrates, such as proteins, carbohydrates and pyruvate (Kengen *et al.*, 1996). It was therefore anticipated

that genes encoding enzymes involved in polymer utilization might be regulated. Preliminary studies revealed the presence of two maltose-regulated genes in *P.furiosus*, *mlrA* and *mlr-2*, the function of which in maltose utilization remains to be elucidated (Robinson *et al.*, 1994; Robinson and Schreier, 1994; Harwoord *et al.*,1997). In this study we have analyzed the regulation of transcription of the *celB* locus of *P.furiosus*, that is involved in the growth on ß-linked glucose polymers. This locus is composed of the *celB* gene for a ß-glucosidase and a divergently orientated gene cluster that includes the *lamA* gene for an extracellular ß-1,3-endoglucanase (Gueguen *et al.*, 1997 Voorhorst *et al.*, 1996; Voorhorst *et al.*, 1997). Here we show that the expression the *celB* gene and the *lamA* operon of which is controlled at the transcription initiation level by growth on ß-linked glucose polymers.

Results

Genetic organization of the celB *gene and the* lamA *operon*

Upstream and in opposite orientation of the *P.furiosus celB* gene, coding for a ß-glucosidase (Voorhorst *et al.*, 1995), a cluster of three genes has been found that here has been designated the *lamA* operon (Fig. 1; see below). The *lamA* operon starts with two tandem alcohol dehydrogenase encoding genes, *adhA* and *adhB*, that is directly followed by a gene coding for a ß-1,3-endoglucanase, *lamA* (Fig. 1) (Voorhorst *et al.*, 1997; Gueguen *et al.*, 1997).

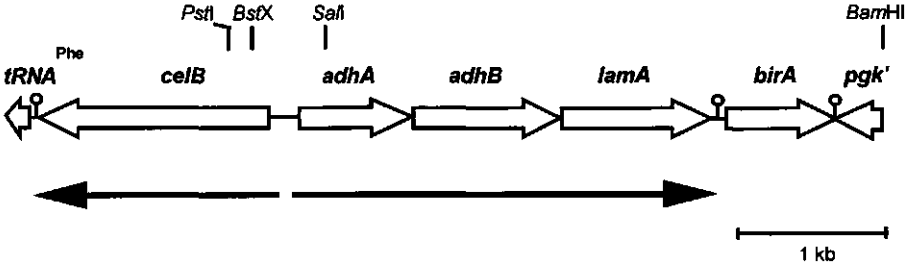


Figure 1. Genetic organization of the *celB* locus. The location and orientation of the genes and transcripts of this locus are indicated with arrows. Relevant restriction sites used for cloning and *in vitro* transcription analysis are indicated.

Downstream of the *lamA* operon an unknown gene is located, designated *birA* based on the high homology (54 % similarity) of its predicted translation product with the biotin ligase domain of BirA-proteins (Howard *et al.*, 1985; Wilson *et al.*, 1992). The *birA* gene is followed by an oppositely orientated open reading frame, which could encode a product with high homology with the C-terminal end of the 2-phosphoglycerate kinases from *Methanothermus fervidus* and *Methanococcus jannaschii*, respectively 63% and 65 % similarity in 157 amino acids (Lehmacher and Hensel, 1994; Bult *et al.*, 1996) (Fig. 1).

The *celB* gene and the *lamA* operon are separated by a 175-bp intergenic region that is highly AT-rich (71%) in comparison with the average AT-content (52%) determined for the pyrococcal genome (Fiala and Stetter, 1986) (Fig. 1; and see below). The stop and start codon of the *adhA-adhB* and *adhB-lamA* genes are separated by only 10 and 12 purine-rich nucleotides, respectively, that may function as ribosome binding sites. The *lamA* gene is followed by a pyrimidine-rich stretch of residues (12 pyrimidines out of 15 residues) with homology to archaeal termination sequences (Brown *et al.*, 1989; Dalgaard and Garrett, 1993; Thomm *et al.*, 1994).While a 41-bp intergenic region separates the *lamA* operon from the first initiation codon of the *birA* gene, the end of the *birA* gene overlapped by 12 nucleotides the *pgk* gene that is located in opposite orientation. In the coding strands, both genes are followed by stretches of pyrimidines residues (mainly thymidines) with homology to archaeal termination sequences.

Transcriptional organization of the* celB *gene and the* lamA *operon

The observed close spacing of the *adhA-adhB-lamA* genes prompted us to study their transcriptional organization and compare this with that of divergent *celB* gene. Northern blot analysis was performed with total RNA extracted from *P.furiosus* cells grown on cellobiose. Hybridization with a ^{32}P-labelled *celB*-specific probe resulted in a hybridizing band with a size of approximately 1.5 kb, which corresponded to the expected size of a monocistronic *celB* transcript (Fig. 2). With a specific probe for the *adhA* gene a hybridizing band of approximatley 2.8 kb could be identified on the Northern blot, which corresponded to the size of a polycistronic transcript that includes *adhA, adhB* and *lamA* (Fig. 2, and see also below). In contrast, a specific probe for the *birA* gene revealed a hybridizing band of approximatley 0.7 kb on the Northern blot (data not shown). Hybridization of the Northern blots with probes specific for the *adhB*

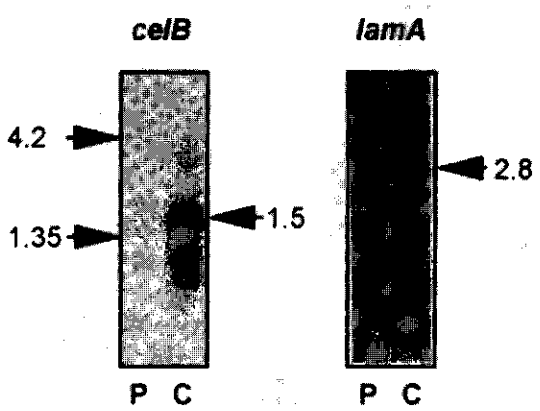


Figure 2. Northern blot analysis of total RNA extracted from *P.furiosus* cells grown on various substrates. *P.furiosus* cells were grown on pyruvate (P),, and cellobiose (C). Specific probes included ^{32}P-labeled fragments of the *celB* and *lamA* genes as indicated.

and *lamA* genes showed the same hybridizing 2.8-kb transcript that was induced under the same conditions (see below; Fig. 5). These results indicate that the *lamA* operon of *P.furiosus* is transcribed as a polycistronic mRNA, including the *adhA, adhB*, and *lamA* genes, and is likely to terminate at the thymidine-rich region downstream of *lamA*.

Common induction of CelB, AdhA and LamA activity

Cells of *P.furiosus* grown on cellobiose as sole carbon and energy source are known to contain high activity levels of the intracellular cellobiose-hydrolysing ß-glucosidase CelB (see below; Kengen *et al.*, 1993). In contrast, pyrococcal cells grown on pyruvate showed only a low level of this activity. Since in these pyruvate-grown cells no CelB protein could be identified by immunological analysis (data not shown) and no *celB* transcript was found by Northern blot analysis (see below), it is likely that the obsered low ß-glucosidase activity is due to other hydrolysing enzymes that are known to be present in *P.furiosus*, such as a ß-mannosidase (Bauer *et al.*, 1996; Kengen *et al.*, 1996).

For the identification of compounds that could induce the CelB activity, cells of *P.furiosus* were grown on pyruvate, maltose, cellobiose, laminarin or peptides and analyzed the presence of CelB activity and protein. High activity levels of CelB (12-18

U/mg) were observed in cell-free extracts of cells grown on cellobiose or laminarin. In contrast, background activity levels of CelB (1.8 ± 0.3 U/mg) were found in cells were grown on pyruvate, maltose or peptides. Similar background levels were observed when *P.furiosus* cells were grown on pyruvate supplemented with either glucose or isopropyl-ß-D-thiogalactopyranoside.

To investigate the induction of the *lamA* operon enzymes, pyrococcal cells were grown on different substrates and analyzed for the presence of the alcohol dehydrogenase AdhA, and the extracellular ß-1,3-endoglucanase LamA. Specific antisera were used for this analysis, since activity measurements were hampered by background activities and residual substrates in cells and supernatant. This immunological detection showed that AdhA and LamA were present in cell-free extracts and culture supernatant, respectively, when pyrococcal cells were grown on cellobiose or laminarin (Fig.3). In contrast, when cells of *P.furiosus* were grown on maltose or pyruvate, neither AdhA nor LamA protein could be detected (Fig.3). These results indicate that enzymes encoded by the *lamA* operon are induced in the same way as the CelB activity by growth of *P.furiosus* on the ß-linked sugars cellobiose and laminarin.

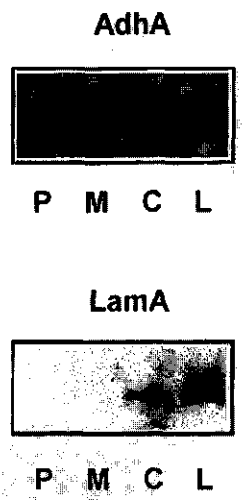


Figure 3. Western blot analysis of AdhA and LamA of *P.furiosus* when grown on different substrates. *P.furiosus* cells were grown on pyruvate (P), cellobiose (C), laminarin (L), and maltose (M). Cell-free extracts (A) and concentrated culture supernatants (B) of pyrococcal cells were separated by SDS-PAGE and AdhA and LamA were detected using polyclonal antiserum against the enzymes purified from *E.coli*.

Regulation and kinetics of induction of the* celB *gene expression

To analyze the control of the CelB activity in *P.furiosus*, Northern blot analyses were performed with total RNA extracted from cells grown on different carbon and energy sources. Hybridization with a *celB*-specific probe showed a high signal intensity with RNA extracted from cells that showed high CelB activity levels, i.e grown on cellobiose or laminarin. No hybridizing bands were detected with RNA extracted from *P.furiosus* cells grown on pyruvate, maltose or peptides (Fig. 2) (data not shown). These results indicate that the *celB* transcription is strongly regulated by the available carbon source.

To study the effect of the ß-linked glucose polymers on the *celB* gene expression in more detail, *P.furiosus* was grown on pyruvate and supplemented with cellobiose at different stages during growth. Specific ß-glucosidase activity was found to increase most rapidly when cellobiose was added to a pyruvate-grown culture in the mid-exponential phase. After addition of cellobiose the specific ß-glucosidase activity rapidly increased within 1h from the background level up to 9 U/mg (Fig. 4A).

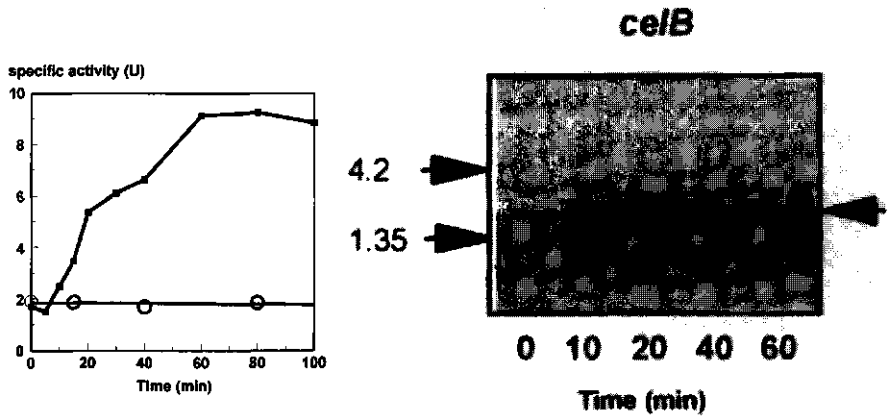


Figure 4. Kinetics of the induction of ß-glucosidase activity and *celB* transcription by cellobiose.
A Specific β-glucosidase activity in *P.furiosus* cells grown on pyruvate following the addition of cellobiose (■-■) or without (0—0). **B.** Northern blot analysis of total RNA extracted from *P.furiosus* cells grown on pyruvate following addition of cellobiose. Hybridization was performed with a ^{32}P-labelled probe specific for *celB*.

To determine whether the observed CelB induction takes place at the transcriptional level, Northern blot analysis was performed on samples from an induction experiment in which pyruvate-grown cells were supplemented with cellobiose at mid-exponential growth phase. A hybridizing band with the expected size of 1.5 kb was obtained with a *celB*-specific probe when RNA was extracted from induced cells,

but not from non-induced, pyruvate-grown cells (Fig. 4B). In the induced cells, which had a doubling time of approximately 50 min, the *celB* transcript was detectable within 10 min after induction with cellobiose and the intensity reached an maximum after 20 min (Fig. 4B).

***Regulation of transcript of the* lamA *operon in* P.furiosus**

The presence of *celB, adhA* and *lamA* gene expression products in pyrococcal cells grown on different substrates suggested a co-regulation of the *celB* gene and the *lamA* operon. To analyze whether the transcription of the *lamA* operon is indeed induced by cellobiose, we performed Northern blot analysis with total RNA extracted from *P.furiosus* cells grown on pyruvate, cellobiose or maltose. Northern blot hybridization with a specific probe for the *adhA* transcript resulted in a band with the expected size with RNA extracted from cells grown on cellobiose, but not with RNA from cells grown on maltose or pyruvate (Fig. 5). Subsequent hybridization with specific probes for *adhB* and *lamA* resulted in the same hybridization pattern as with the *adhA* specific probe (Fig. 5). Moreover, the transcript that hybridized to all probes showed the same size as expected from the operon structure (see also Fig. 1).

Transcription initiation of the* lamA *operon and* celB *gene and cell-free transcription

For the identification of the transcription initiation sites of the two divergent transcripts, primer extension experiments were performed using total RNA extracted from *P.furiosus*. The transcription initiation start site of the *lamA* transcript could be identified 10 nucleotides upstream of the translation start site, using total RNA extracted from *P.furiosus* cells grown on cellobiose (Fig. 6A). At a conserved distance of 25 nucleotides, the transcription start site was preceeded by a hexanucleotide sequence that resembles the archaeal TATA box sequence (Fig. 7) (Brown *et al.*, 1989; Dalgaard and Garrett, 1993; Thomm, 1996).

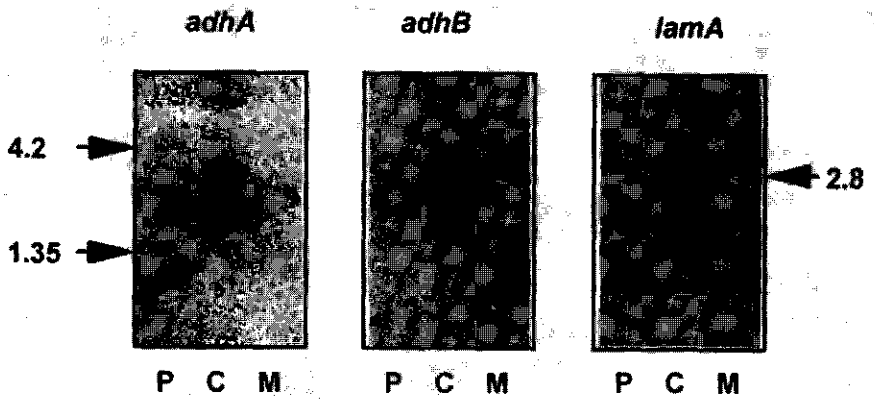


Figure 5. Detection of the transcript of the *lamA* operon in RNA extracted from cells grown on pyruvate (P), cellobiose (C), and maltose (M). The Northern blots were hybridized with the indicated ^{32}P-labeled probes.

A cell-free transcription system for *Pyrococcus*, that consists of TBP, TFIIB and RNA polymerase (Hausner *et al.*, 1996), efficiently initiates transcription from the promoter of the *P.furiosus gdh* gene, coding for a glutamate dehydrogenase (Eggen *et al.*, 1993; Hethke *et al.*, 1996). This system has been applied to analyze transcription initiation of the divergently orientated *lamA* operon (Fig. 6A) and *celB* gene (Fig. 6B). Primer extension on the *in vitro* produced transcript of the *celB* gene and the *lamA* operon showed that initiation *in vitro* occured at the same site as *in vivo* (Fig. 6B) (Voorhorst *et al.*, 1995). To compare the *in vitro* activity of these regulated promoters with the strong constitutively expressed *gdh* promoter, RNA products synthesized at linearized templates were analysed in run-off transcription experiments. When plasmid pLUW504 (Voorhorst *et al.*, 1995) was cleaved in the *celB* coding region with *Bst*XI or in the *adhA* coding region with *Sal*I (Fig. 1), run-off transcripts of 185 and 244 nucleotides should be synthesized. Transcripts of this size were observed as expected (Fig. 6C, lanes 2 and 3). Quantitation of the run-off products revealed that the activity of the *celB* promoter was only 1/90, and that of the *adhA* promoter only 1/600 of that of the *gdh* promoter obtained with template pLUW479 (Hetke *et al.*, 1996).

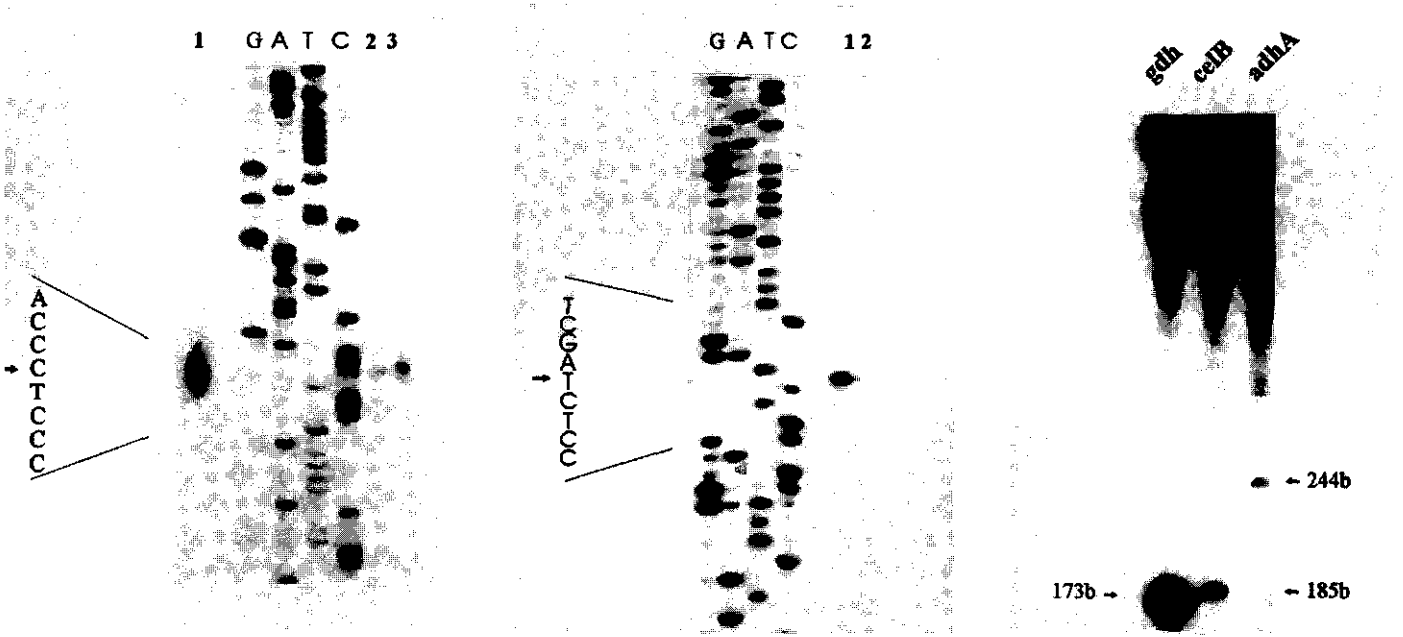


Figure 6. *In vivo* and *in vitro* transcriptional analysis of the *celB* locus. **A.** Analysis of the transcription initiation site of the *lamA* operon at the *adhA* promoter by primer extension. Primer extended products on *in vivo* (lane 1) and *in vitro* (lane 2) produced transcripts were electrophoresed in parallel with the sequence ladder (lanes G, A, T, C) generated with the same primer on the non-coding strand of *adhA*. Annealing reactions were performed at 60°C (lane 2) and 50°C (lane 3). **B.** Analysis of the transcription initiation site at the *celB* promoter *in vitro* by primer extension. Primer extended products were electrophoresed in parallel with the sequence ladder (lanes G, A, T, C) generated with the same primer on the non-coding strand of *celB*. Annealing reactions were performed at 50°C (lane 1) and 60°C (lane 2). **C.** Comparasion of the *in vitro* transcription efficiency of the *gdh*, *celB* and *adhA* genes. RNA polymerase, the Superdex fraction of a TFB as well as the heparin-Sepharose fraction of a TFA were assayed for specific activity in cell-free transcription system reactions as described in (Hethke *et al.*, 1996). Labelled run-off transcripts of the *gdh* (Hethke *et al.*, 1996), *celB* and *adhA* genes were separated on a 8% polyacrylamide-urea gel and identified by autoradiography. Only one fifth of the product generated with the *gdh* template was loaded. The arrows label the run-off transcripts and show their length in nucleotides.

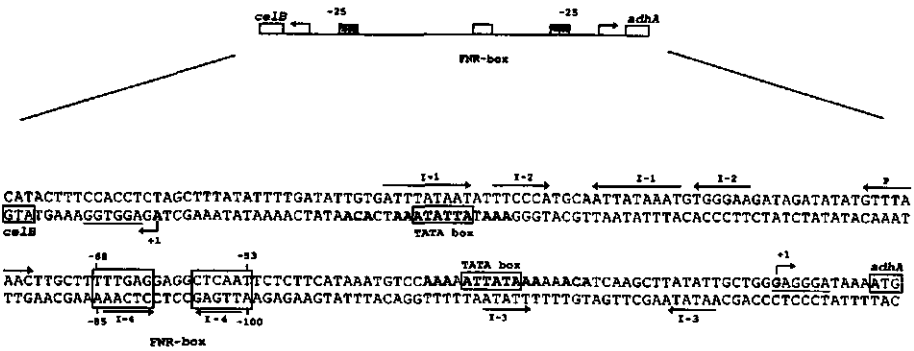


Figure 7. The intergenic region between the *celB* and the *adhA* genes. Transcription initiation sites are indicated by arrows (+1); conserved residues surrounding the TATA boxes (boxed) are indicated in bold face, the putative ribosome binding sites are underlined and initiation codons are boxed. For clarity reasons only inverted repeats (I-1 to I-4) and the palindromic sequence (P) are shown.

Discussion

To provide insight in the way gene expression is controlled in hyperthermophilic archaea, we studied the transcriptional control of the *P.furiosus celB* gene and the divergently transcribed *lamA* operon. The transcription initiation of the *celB* gene and the *lamA* operon of *P.furiosus* was found to be controlled by the ß-linked glucose polymers, cellobiose and laminarin. The *celB* gene encodes a ß-glucosidase, and the *lamA* operon includes three clustered genes, the last of which, *lamA*, encodes the second glycosyl hydrolase of the locus, the extracellular ß-1,3-endoglucanase. The *celB* gene is transcribed as monocistronic mRNA indicating the functionality of the typical archaeal termination sequence downstream of its coding region (Fig. 1) (Voorhorst *et al.*, 1995). Northern blot analysis showed that the transcript of the *lamA* operon is a polycistronic mRNA. Termination of this 2.8 kb transcript most likely occurs downstream of the *lamA* gene at a stretch of pyrimidine-rich residues, since the following *birA* gene is transcribed in a single mRNA that is probably inititated in the 41-bp intergenic region.

Only few reports exist to date on the control of archaeal gene expression. The identification of the divergent transcripts of the *P.furiosus celB* locus allowed for a more detailed analysis of transcriptional regulation in hyperthermophilic *Archaea*. Induction studies and their kinetics showed a rapid increase of ß-glucosidase activity levels upon

addition of ß-linked glucose polymers to pyruvate growing cells. This was corrobated by transcriptional analysis that revealed a tight control at transcriptional level of the *celB* expression in *P.furiosus* and a rapid induction response within one fifth of the generation time in this hyperthermophilic archaeon.

Several of the enzymes encoded by genes of the *celB* locus have been implied to function cooperatively in substrate utilization of the large glucose polymer laminarin. The extracellular ß-1,3-endoglucanase LamA is involved in the hydrolysis of this polymer into glucose oligomers that, following transport, are intracellularly degraded by the β-glucosidase CelB to generate glucose (Gueguen *et al.*, 1997). The resulting glucose can be further metabolized via the modified Embden-Meyerhof-Parnas pathway that operates in *P.furiosus* (Kengen *et al.*, 1994). In agreement with the key role of the two glycosyl hydrolases, high activities of the intracellular ß-glucosidase and large amounts of the extracellular ß-1,3-endoglucanase were detected in cultures grown on cellobiose or laminarin. No or only background levels of these enzymes were observed in pyrococcal cells grown on pyruvate or maltose. In accordance with the *celB* and the *lamA* transcripts were only found with total RNA extracted from cells grown on cellobiose or laminarin, indicating co-regulation of the divergently orientated transcripts.

Initiation of transcription of the *celB* gene and the *lamA* operon using the cell-free transcription system for *P.furiosus* occured at the same sites as identified *in vivo*. Nevertheless, the efficiency of transcription initiation of these transcripts in the cell-free system was more than 90-fold lower than found with the *gdh* gene that is known to be expressed efficiently in *P.furiosus* (Hethke *et al.*, 1996). This may indicate that an activator is required for efficient transcription initiation of these genes in *P.furiosus*.

The archaeal TATA box, equivalent to the eucaryal TATA-box, has been shown to interact with the TBP and is involved in directing the transcription (Rowlands *et al.*, 1994; Hausner *et al.*, 1996). Remarkably, the promoters of the *celB* gene and the *lamA* operon contain two identical TATA box sequences (ATTATA) both spaced by 25 nucleotides from their transcription initiation site. Moreover, the sequences surrounding the TATA box sequences of both genes showed high degree of conservation (Fig. 7) and may play a role in determining the efficiencty of transcription intitiation. The upstream region of the TATA box sequence is expected to interact with the initiation complex, in which the TFIIB stabilizes the TBP-DNA binary comples (Hausner *et al.*, 1996). The highly A-T rich (71 %) intergenic region was further analyzed for features that could be related to the control of the transcriptional co-regulation of the *celB* gene

and the *lamA* operon and was found to consist of four inverted and direct repeats of more than 5 residues, and a palindromic sequence (Fig. 7). An almost perfect inverted repeat (labeled I-4) is present with a striking similarity with the binding-sites for the FNR/CRP family of bacterial transcription regulators, and may be involved in the coordinated regulation of the divergently orientated *celB* gene and *lamA* operon. The location of this putative binding site (-54 to -68 from the *adhA* and -86 to -100 from the *celB* transcription initiation sites) is in agreement with that found for activators in bacterial systems (Gralla and Collado-Vides, 1996). These results suggest the presence of a bacterial-like activator system that should control gene expression in the hyperthermophilic archaeon *P.furiosus*.

Experimental procedures

Sequence analysis of the celB *locus of* P.furiosus

To complete the nucleotide sequence of the *P.furiosus celB* locus we sequenced the pyrococcal DNA in pLUW500 (Voorhorst *et al.*, 1995) covering the 3.5-kb *Pst*I-*Bam*HI fragment, including the *adhA, adhB* and *lamA* genes (Fig. 1). Nucleotide sequence was carried out on automated DNA sequencers Applied Biosystems 373A using a Prism™ Ready Reaction DyeDeoxy™ Terminator cycle sequencing kit or the Li-Cor 4000L with the Thermo Sequenase fluorescent labelled primer cycle sequencing kit with 7-deaza-dGTP (Amersham) using IR-labelled oligonucleotides (MWG-Biotech).

Computer analysis of nucleotide and deduced amino acid sequences were carried out with the PC/GENE program version 5.01 (IntelliGenetics Inc.) and the GCG package version 7.0 (Devereux *et al.*, 1990) at the CAOS/CAMM centre of the University of Nijmegen (The Netherlands) and the multiple sequence program ClustalW 1.6 available on internet (BCM Search Launcher, Human Genome Center, Baylor College of Medicine, Houston TX).

Growth and induction conditions

P.furiosus strain DSM3638 was cultured at 98°C in synthetic seawater as previously described (Kengen *et al.*, 1993) using pyruvate (40 mM), cellobiose (10 mM), maltose (10 mM) or laminarin (2 g/l) as growth substrate. Cells were grown for several generations on these substrate to allow adaptation before any analysis was performed. Growth of *P.furiosus* was followed by spectrophotometrically analyzing the increase in

optical density and determining the hydrogen production using a Packard gaschromatograph. Under these conditions late exponential growth phase corresponds with an OD at 450 nm of 0.8-0.9.

For the induction with cellobiose, cells were grown on pyruvate to OD at 450 nm of 0.7-0.8 and supplemented with cellobiose (10 mM). Samples were taken before and after induction and immediately cooled on ice prior to further processing.

Preparation of cell-free extracts and enzyme activities

Cells grown to late exponential growth phase were harvested by centrifugation, washed with fresh medium, and subsequently resuspended in citrate buffer as described previously (Voorhorst *et al.*, 1995). Following sonication to disrupt the cells, the cell debris was pelleted and the resulting supernatant was used as cell-free extract for activity measurements and immunological analysis. The ß-glucosidase activity was determined using by hydrolysis of ß-D-glucopyranoside-*p*-nitrophenyl (Boehringer, Mannheim GmbH, FRG) (Kengen *et al.*, 1993). Protein concentration was measured using Bradford reagents (Bio-Rad Laboratories) with BSA as standard.

Polyacryl amide gel electrophoresis and Western-blotting

Electrophoretic analysis of protein samples was performed by an 11% sodium dodecyl sulphate-polyacrylamide gel electrophoresis (SDS-PAGE) (Laemmli, 1970). Protein samples for SDS-PAGE were prepared by heating for 5 min at 100°C in an equal volume of sample buffer (0.1 M Tris-HCl, 5% SDS, 0.9% 2-mercaptoethanol, 20% glycerol, pH 6.8). Proteins separated by SDS-PAGE were transferred to nitrocellulose membrane by semi-dry blotting (Bio-Rad Laboratories). Immunological detection was performed as previously described (Voorhorst *et al.*, 1997) using antiserum raised against the AdhA and LamA protein purified from *E.coli* (Gueguen *et al.*, 1997; Voorhorst *et al.*, 1997).

Isolation of total mRNA, Northern analysis and primer extension

Cells were harvested by centrifugation and RNA was isolated using guanidinium isothyiocyanate and ß-mercaptoethanol as previously described (Voorhorst *et al.*, 1995). For Northern blot analysis 15 µg RNA was separated on a formaldehyde-1% agarose gel. Following gel electrophoresis the RNA was transferred by capillary blotting to a Hybond N^+-membrane (Sambrook *et al.*, 1989). Gene-specific probes were obtained after appropriate digestion of DNA from the plasmid pLUW500 or pLUW501

(Voorhorst *et al.*, 1995), purified by Gene Clean (Bio 101, La Jolla, CA) and labelled by nick-translation (Sambrook *et al.*, 1989). During the hybridizations Southern blots were included to verify the specificity of the probes. Primer extension experiments on the isolated or synthesized RNA templates were done as previously described (Hethke *et al.*, 1996). The following oligonucleotides were used for the indicated templates 5'-CCA AGA ATA TCC AAA CAT GAA G-3' (*celB*) and 5'-GGC AAT CTT CTC TAA CCT ATC AAC-3' (*adhA*)

***Cell-free transcription system of* P.furiosus.**

Cell-free transcription reactions using partially purified transcription factors and highly purified RNA polymerase (Superdex fraction) were essentially done as previously described (Hethke *et al.*, 1996). Cell-free transcription assays were performed at 70°C with a potassium chloride concentration of 300 mM.

Nucleotide sequence accession number.

The nucleotide sequence reported has been submitted to GenBank/EMBL Data Bank with the accesion number xxxxx

Acknowledgement

We are grateful to Ans Geerling for assistance with determination of the nucleotide sequences. Part of this work was supported by contracts BIOT-CT93-0274 and BIOT-CT96-0488 of the European Union

References

Bauer, M.W., Bylina, E.J., Swanson, R.V., and Kelly, R.M. (1996) Comparison of a β-glucosidase and a β-mannosidase from the hyperthermophilic archaeon *Pyrococcus furiosus*. Purification, characterization, gene cloning, and sequence analysis. *J. Biol. Chem.* **271**: 23749-23755.

Baumann, P., Qureshi, S.A., and Jackson, S.P. (1995) Transcription: new insights from studies on *Archaea*. *Trends Genet.* **11**: 279-283.

Brown, J.W., Daniels, C.J., and Reeve, J.N. (1989) Gene structure, organization, and expression in archaebacteria. *CRC Critic. Rev. Microbiol.* **16**: 287-338.

Bult, C.J., White, O., Olsen, G.J., Zhou, L., Fleischmann, R.D., Sutton, G.G., Blake, J.A., FitzGerald, L.M., Clayton, R.A., Gocayne, J.D., Kerlavage, A.R., *et al.* (1996) Complete genome sequence of the methanogenic archaeaon, *Methanococcus jannaschii*. *Science* **273**: 1058-1073.

Ciaramella, M., Cannio, R., Moracci, M., Pisani, F.M., and Rossi, M. (1995) Molecular biology of extremophiles. *World J. Microbiol. Biotechnol.* **11**: 71-84.

Cohen-Kupiec, R., Blank, C., and Leigh, J.A. (1997) Transcriptional regulation in Archaea: in vivo demonstration of a repressor binding site in a methanogen. *Proc. Natl. Acad. Sci. USA* **94**: 1316-1320.

Dalgaard, J.Z., and Garrett, R.A. (1993) Archaeal hyperthermophile genes. In *The biochemistry of Archaea (Archaebacteria).* M. Kates, Kusher, D.J., and Matheson, A.T. (eds). Amsterdam: Elsevier. pp. 535-563.

Dedecker, B.S., O'Brien, R., Fleming, P.L., Geiger, J.H., Jackson, S.P., and Sigler, P.B. (1996) The crystal structure of a hyperthermophilic archaeal TATA-box binding protein. *J. Mol. Biol.* **264**: 1072-1084.

Eggen, R.I.L., Geerling, A.C.M., Waldkötter, K., Antranikian, G., and de Vos, W.M. (1993) The glutamate dehydrogenase-encoding gene of the hyperthermophilic archaeon *Pyrococcus furiosus*: sequence, transcription and analysis of the deduced amino acid sequence. *Gene* **132**: 143-148.

Fiala, G., and Stetter, K.O. (1986) *Pyrococcus furiosus* sp. nov. represents a novel genus of marine heterotrophic archaebacteria growing optimally at 100°C. *Arch. Microbiol.* **145**: 56-61.

Fitz-Gibbon, S., Choi, A.J., Miller, J.H., Stetter, K.O., Simon, M.I., Swanson, R., and Kim, U. (1997) A fosmid-based genomic map and identification of 474 genes of the hyperthermophilic archaeon *Pyrobaculum aerophilum. Extremophiles* 1: 36-51.

Frey, G., Thomm, M., Brudigam, B., Gohl, H.P., and Hausner, W. (1990) An archaebacterial cell-free transcription system. The expression of tRNA genes from *Methanococcus vannielii* is mediated by transcription factor. *Nucleic Acids Res.* **18**: 1361-1367.

Gralla, J.D., and Collado-Vides, J. (1996) Organization and function of transcription regulatory elements. In *Escherichia coli and Salmonella cellular and molecular biology.* F. C. Neidhardt (Ed. in Chief). Washington: ASM Press. second edition. pp.1232-1262.

Gueguen, Y., Voorhorst, W.G.B., van der Oost, J., and de Vos, W.M. (1997) Molecular and biochemical characterization of an endo-β-1,3-glucanase of the hyperthermophilic archaeon *Pyrococcus furiosus. J. Biol. Chem.* (in press)

Harwood, V.J. Denson, J.D., Robinson-Bidle, K.A., Schreier, H.J. (1997) Overexpression and characterization of a prolyl endopeptidase from the hyperthermophilic archaeon *Pyrococcus furiosus. J. Bacteriol.* **179**: 3613-3618.

Hausner, W., Frey, G., and Thomm, M. (1991) Control regions of an archaeal gene. A TATA box and an initiator element promote cell-free transcription of the $tRNA^{Val}$ gne of *Methanococcus vannielii. J. Mol. Biol.* **222**: 495-508.

Hausner, W., Wettach, J., Hethke, C., and Thomm, M. (1996) Two transcription factors related with the eucaryal transcription factors TATA-binding protein and transcription factore IIB direct promoter recognition by an archaeal RNA polymerase. *J. Biol. Chem.* **271**: 30144-30148.

Hensel, R. (1993) Proteins of extreme thermophiles. In *The biochemistry of Archaea (Archaebacteria).* M. Kates, Kushner, D.J., and Matheson, A.T. (eds). Amsterdam: Elsevier. pp. 209-221.

Hethke, C., Geerling, A.C.M., Grondahl, B., de Vos, W.M., and Thomm, M. (1996) A cell-free transcription system for the hyperthermophilic archaeon *Pyrococcus furiosus. Nucleic Acids Res.* **24**: 2369-2376.

Hethke, C., and Thomm, M. (1997) Unpublished data.

Howard, P.K., Shaw, J., and Otsuka, A.J. (1985) Nucleotide sequence of the *birA* gene encoding the biotin operon repressor and biotin holoenzyme synthetase functions of *Escherichia coli. Gene* **35**: 321-331.

Hüdepohl, U., Reiter, W.D., and Zillig, W. (1990) In vitro transcription of two rRNA genes of the archaebacterium *Sulfolobus* sp. B 12 indicates a factor requirement for specific initiation. *Proc. Natl. Acad. Sci. USA* **87**: 5851-5855.

Ken, R., and Hackett, N.R. (1991) *Halobacterium* halobium strains lysogenic from phage ΦH contain a protein resembling coliphage repressors *J. Bacteriol.* **173**: 955-960.

Kengen, S.W.M., de Bok, F.A.M., van Loo, N.D., Dijkema, C., Stams, A.J.M., and de Vos, W.M. (1994) Evidence for the operation of a novel Embden-Meyerhof pathway that involves ADP-dependent kinases during sugar fermentation by *Pyrococcus furiosus*. *J. Biol. Chem.* **269**: 17537-17541.

Kengen, S.W.M., Luesink, E.J., Stams, A.J.M., and Zehnder, A.J.B. (1993) Purification and characterization of an extremely stable β-glucosidase from the hyperthermophilic archaeon *Pyrococcus furiosus*. *Eur. J. Biochem.* **213**: 305-312.

Kengen, S.W.M., Stams, A.J.M., and de Vos, W.M. (1996) Sugar metabolism of hyperthermophiles. *FEMS Micrbiol. Rev.* **18**: 119-137.

Laemmli, U.K. (1970) Cleavage of structural proteins during the assembly of the head of bacteriophage T4. *Nature* **227**: 680-685.

Langer, D., Hain, J., Thuriaux, P., and Zillig, W. (1995) Transcription in *Archaea*: Similarity to that of *Eucarya*. *Proc. Natl. Acad. Sci. USA* **92**: 5768-5772.

Lehmacher, A., and Hensel, R. (1994) Cloning, sequencing and expression of the gene encoding 2-phosphoglycerate kinase from Methanothermus fervidus. *Mol. Gen. Genet.* **242**: 163-168.

Morgan, R.M., Pihl, T.D., Nölling, J., and Reeve, J.N. (1997) Hydrogen regulation of growth, growth yields, and methane gene transcription in *Methanobacterium thermoautotrophicum* ΔH. *J. Bacteriol.* **179**: 889-898.

Nölling, J., and Reeve, J.N. (1997) Growth- and substrate-dependent transcription of the fomate dehydrogenase (*fdhCAB*) operon in *Methanobacterium thermoformicicum* Z-245. *J. Bacteriol.* **179**: 899-908.

Palmer, J.R., and Daniels, C.J. (1995) In vivo definition of an archaeal promoter. *J. Bacteriol.* **177**: 1844-1849.

Reiter, W.D., Zillig, W., and Palm, P. (1988) In: *Advances in Virus Research*, Vol. 34 (Maramorosch, K., Murhy, F.A., and Shatkin, A.J., Eds.), Academic Press, Orlando, FL., pp. 143-188.

Reiter, W.D., Hüdepohl, U., and Zillig, W. (1990) Mutational analysis of an archaebacterial promoter: Essential role of a TATA box transcription efficiency and start-site selection. *Proc. Natl. Acad. Sci. USA* **87**: 9509-9513.

Robinson, K.A., Robb, F.T., and Schreier, H.J. (1994) Isolation of maltose-regulated genes from the hyperthermophilic archaeum, *Pyrococcus furiosus*, by subtractive hybridization. *Gene* **148**: 137-141.

Robinson, K.A., and Schreier, H.J. (1994) Isolation, sequence and characterization of the maltose-regulated *mlrA* gene from the hyperthermophilic archaeum *Pyrococcus furiosus*. *Gene* **151**: 173-176.

Rowlands, T., Baumann, P., and Jackson, S.P. (1994) The TATA-binding protein: A general transcription factor in eukaryotes and archaebacteria. *Science* **264**: 1326-1329.

Sambrook, J., Fritsch, E.F., and Maniatis, T. (1989). *Molecular Cloning. A Laboratory Manual.* Cold Spring Harbor, New York, Cold Spring Harbor Laboratory Press.

Stolt, P., and Zillig, W. (1992) *In vivo* studies on the effects of immunity genes on early lytic transcription in the *Halobacterium salinarium* phage Φ H. *Mol. Gen. Genet.* **235**: 197-204.

Thomm, M. (1996) Archaeal transcription factors and their role in transcription initiation. *FEMS Microbiol. Rev.* **18**: 159-171.

Thomm, M., Hausner, W., and Hethke, C. (1994) Transcription Factors and Termination of Transcription in Methanococcus. *Syst. Appl. Microbiol.* **16**: 648-655.

Voorhorst, W.G.B., Eggen, R.I.L., Luesink, E.J., and de Vos, W.M. (1995) Characterization of the *celB* gene coding for β-glucosidase from the hyperthermophilic archaeon *Pyrococcus furiosus* and its expression and site-directed mutation in *Escherichia coli*. *J. Bacteriol.* **177**: 7105-7111.

Voorhorst, W.G.B., Wittenhorst, V., van der Oost, J., and de Vos, W.M. (1997) Genetic and biochemical characterization of a short-chain and an iron-containing alcohol dehydrogenase from the hyperthermophilic archaeon *Pyrococcus furiosus*. *Submitted for publication.*

Wettach, J., Gohl, H.P., Tschochner, H. and Thomm, M. (1995) Functional interaction of yeast and human TATA-binding proteins with an archaeal RNA polymerase and promoter. *Proc. Natl. Acad. Sci. USA* **92**: 472-476.

Wilson, K.P., Shewchuk, L.M., Brennan, R.G., Otsuka, A.J., and Matthews, B.W. (1992) *Escherichia coli* biotin holoenzyme synthetase/bio repressor crystal structure delineates the biotin- and DNA-binding domains. *Proc. Natl. Acad. Sci. USA* **89**: 9257-9261.

Zillig, W., Palm, P., Klenk, H.P., Langer, D., Hüdepohl, U., Hain, J., Lanzendorfer, M., and Holz, I. (1993) Transcription in Archaea. In *The Biochemistry of Archaea (Archaebacteria).* M. Kates, Kushner, D.J., and Matheson, A.T. (eds). Amsterdam: Elsevier. pp. 367-391.

Zillig, W., Reiter, W-D., Palm, P., Gropp, F., Neumann, H., and Rettenberger, M. (1999) Viruses in archaebacteria. In *The Bacteriophages,* (R.Calender, Ed.), Vol. 1 of The Virusses (Fränkel-Conrat, H. and Wagner, R.R., series Eds.), New York. Plenum Press. pp. 517-558.

Chapter 6

Isolation and Characterization of the Hyperthermostable Serine Protease, Pyrolysin, and Its Gene from the Hyperthermophilic Archaeon *Pyrococcus furiosus*

Wilfried G.B. Voorhorst, Rik I.L. Eggen, Ans, C.M. Geerling, Christ Platteeuw, Roland, J. Siezen, and Willem M. de Vos

reprinted with permission from the **Journal of Biological Chemistry**, 1996, 271 (34), 20426-20431

THE JOURNAL OF BIOLOGICAL CHEMISTRY
Vol. 271, No. 34, Issue of August 23, pp. 20426–20431, 1996
Printed in U.S.A.

Isolation and Characterization of the Hyperthermostable Serine Protease, Pyrolysin, and Its Gene from the Hyperthermophilic Archaeon *Pyrococcus furiosus**

(Received for publication, February 28, 1996, and in revised form, April 26, 1996)

Wilfried G. B. Voorhorst‡, Rik I. L. Eggen‡§, Ans C. M. Geerling‡, Christ Platteeuw‡¶, Roland J. Siezen¶, and Willem M. de Vos‡¶‖

From the ‡Department of Microbiology, Wageningen Agricultural University, 6703 CT Wageningen, and the ¶Department of Biophysical Chemistry, NIZO Ede, The Netherlands

The hyperthermostable serine protease pyrolysin from the hyperthermophilic archaeon *Pyrococcus furiosus* was purified from membrane fractions. Two proteolytically active fractions were obtained, designated high (HMW) and low (LMW) molecular weight pyrolysin, that showed immunological cross-reaction and identical NH_2-terminal sequences in which the third residue could be glycosylated. The HMW pyrolysin showed a subunit mass of 150 kDa after acid denaturation. Incubation of HMW pyrolysin at 95 °C resulted in the formation of LMW pyrolysin, probably as a consequence of COOH-terminal autoproteolysis. The 4194-base pair *pls* gene encoding pyrolysin was isolated and characterized, and its transcription initiation site was identified. The deduced pyrolysin sequence indicated a prepro-enzyme organization, with a 1249-residue mature protein composed of an NH_2-terminal catalytic domain with considerable homology to subtilisin-like serine proteases and a COOH-terminal domain that contained most of the 32 possible *N*-glycosylation sites. The archaeal pyrolysin showed highest homology with eucaryal tripeptidyl peptidases II on the amino acid level but a different cleavage specificity as shown by its endopeptidase activity toward caseins, casein fragments including α_{S1}-casein and synthetic peptides.

Hyperthermophilic microorganisms with an optimum growth temperature of at least 80 °C have been studied in recent years to gain insight in biochemical adaptations to their extreme environment (1, 2). While few hyperthermophiles utilize carbohydrates, which in some cases are fermented via novel pathways (3, 4), most of these unusual microorganisms are capable of growing rapidly on proteins and peptides (2, 5). Several extracellular proteases that may be involved in the first step of protein utilization have been characterized, and all are derived from hyperthermophilic microorganisms that belong to the domain of the Archaea, the third lineage of life (6). A cell envelope-associated protease from *Sulfolobus acidocaldarius* has been purified and shown to belong to a new class of acid proteases based on its gene sequence that showed the presence of a signal sequence and several potential *N*-glycosylation sites (7, 8). Other hyperthermostable proteases that have been biochemically characterized include a thiol protease from *Pyrococcus* (9) and serine proteases from *Desulfurococcus* (10), *Thermococcus stetteri* (11), and *Pyrococcus furiosus* (12–15). Moreover, a serine protease gene has been described in *Pyrobaculum aerophilum*, but its product has not yet been characterized (16).

All purified archaeal proteases show a high thermostability and thermoactivity, but the most thermostable protease to date is pyrolysin from *P. furiosus* with a half-life value of 4 h at the normal boiling point of water (12, 13). Here we describe the purification, post-translational modification, and substrate specificity of pyrolysin. Via reversed genetics the corresponding *pls* gene has been cloned, and its transcription initiation and nucleotide sequence were determined. The results indicate that pyrolysin has a prepro-enzyme structure and is a true endopeptidase belonging to the family of subtilisin-like serine proteases (17).

MATERIALS AND METHODS

Bacterial Strains, Plasmids, and Culture Conditions—P. furiosus (DSM3638) cells were cultivated at 95 °C in a 30-liter stainless steel fermentor (Bioengineering, Wald, Switzerland) with a 20-liter working volume, sparged with N_2 gas. The growth medium consisted of synthetic seawater containing proteinaceous material (peptone 5 g/liter, yeast extract 1 g/liter) without elemental sulfur as described previously (12). *Escherichia coli* strain TG1 (18) was grown in L broth. The vector pUC19 (Pharmacia Biotech Inc.) was used for cloning and sequencing procedures.

Purification and Characterization of Pyrolysin—P. furiosus cells (35 g) were resuspended in 25 ml of P-buffer (50 mM sodium phosphate, pH 6.5), homogenized by sonication, and passed through a French press. The membrane fraction was separated from the cytoplasmic fraction by centrifugation (17,500 × *g*, 20 min), resuspended in 10 ml of P-buffer supplemented with 6 M urea, and incubated for 4 h at 95 °C. After centrifugation to remove the cell debris and denatured proteins, sample buffer (6 M urea, 0.1 M Tris-HCl, 0.9% 1-mercaptoethanol, 20% glycerol, pH 6.8) was added to the supernatant that was subjected to preparative urea gel electrophoresis (urea-PAGE)[1] with the Bio-Rad model 491 Prep Cell (Bio-Rad). The gel (11% acrylamide, 10 × 2.5 cm) was made up to 6 M urea, and proteins were electroeluted in native PAGE-buffer (192 mM glycine, 25 mM Tris, pH 8.5) and analyzed for activity. Fractions containing proteolytic activity were pooled in a high molecular weight (HMW pyrolysin) and a low molecular weight fraction (LMW pyrolysin) and loaded onto an anion exchange column (1 × 3 cm) of Fractogel® EMD trimethylaminoethyl-650 (M) (Merck) equilibrated in 50 mM Tris-HCl, pH 8.5. Subsequently, the buffer system of the anion-exchange column was changed to P-buffer, and after equilibration, the bound

* Part of this work was supported by contract BIOT-CT93–0274 of the European Union. The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

The nucleotide sequence(s) reported in this paper has been submitted to the GenBank™/EBI Data Bank with accession number(s) U55835.

§ Current address: EAWAG, 8600 CH, Dübendorf, Switzerland.

‖ To whom correspondence should be addressed: Hesselink van Suchtelenweg 4, 6703 CT Wageningen, The Netherlands. Tel.: 31-3174-83100; Fax: 31-3174-83829; E-mail: willem.devos@algemeen.micr.wau.nl.

[1] The abbreviations used are: PAGE, polyacrylamide gel electrophoresis; HPLC, high pressure liquid chromatography; TPPs, tripeptidyl peptidases II; ORF, open reading frame; kb, kilobase pair(s).

proteins were eluted by a linear gradient 0–1.5 M NaCl in P-buffer. Fractions containing proteolytic activity were pooled, dialyzed, and concentrated using a diaflow cell with a cut-off value of 30 kDa (Filtron Technology Corp.). The NH_2-terminal amino acid sequences of the purified proteases were determined, either from immobilized proteins that had been separated by SDS-PAGE or directly from a purified preparation, by Edman degradation using an Applied Biosystems 477A Protein Sequencer (Applied Biosystems). Purified protease fractions were used to raise antibodies in rabbits, and crude serum was used in immunoblot experiments as described previously (19). Glycoproteins were stained following urea-SDS-PAGE by the periodic acid-Schiff technique using fuchsin-sulfite stain (Sigma-Aldrich) (20).

Polyacrylamide Gel Electrophoresis—SDS-PAGE was carried out according to Laemmli (21) using 9.5% polyacrylamide gels. Acid denaturation of pyrolysin prior to electrophoresis was achieved by incubation in 5 M formic acid for 45 min at room temperature. Equimolar amounts of potassium hydroxide were used for subsequent neutralization. Analytical urea-SDS-PAGE was essentially the same as SDS-PAGE with the addition of 6 M urea, and samples were prepared as described for the preparative urea-PAGE using sample buffer supplemented with 5% SDS. A method was developed for the *in situ* detection of proteolytic active proteins on urea-(SDS)-PAGE using an overlayer of Kodak X-OMAT AR röntgen film (Eastman Kodak Co.). Following a 10-min incubation at 95 °C the gelatin layer of the Kodak film was transferred to the gel and degraded on places where proteolytic activity was present. The nondegraded gelatin background was subsequently stained with Coomassie Brilliant Blue, resulting in a blue gel with cleared bands of activity.

Activity Assays and Substrate Specificity—The proteolytic activity of pyrolysin toward casein was assayed for 30 min at 95 °C as described previously (12). Total casein and purified α_{S1}-, β-, and k-casein were incubated with pyrolysin, and degradation products were analyzed by SDS-PAGE (22). The enzyme activity toward synthetic substrates (1 mM final concentration) was assayed in 0.5 ml of 0.1 M potassium phosphate buffer, pH 7.5, at 95 °C and followed spectrophotometrically at 405 nm on a Beckman DU7500 spectrophotometer (Beckman Instruments Inc.). Synthetic substrates were obtained from Bachem Feinchemikalien (Bubendorf, Switzerland) or from Chromogenix AB (Mölndal, Sweden). The tripeptidyl-peptidase substrate Ala-Ala-Phe-*p*-nitroanilide and its inhibitor, Arg-Ala-δAla-Val-Ala (where δAla is δAla, dehydroalanine), were kindly provided by B. Tomkinson (University of Uppsala, Sweden) (23, 24). The cleavage specificity of pyrolysin in the peptide α_{S1}-casein-(1–23) was determined by HPLC analysis of initial cleavage products (22, 25) obtained by incubating 25 ml of a solution of α_{S1}-casein-(1–23) (10 mg/ml), 1 ml of the purified LMW pyrolysin (0.25 mg/ml), and 50 ml of 0.1 M potassium phosphate buffer, pH 7.5, at 95 °C. At various times, aliquots of 20 ml were cooled and supplemented with an equal volume of solvent B, composed of acetonitrile/water/trifluoroacetic acid (900:100:0.7, v/v). As a reference the degradation pattern of α_{S1}-casein-(1–23) obtained by the PIII-type protease from *Lactococcus lactis* was used as described previously (22, 25).

Cloning of the pls Gene Encoding Pyrolysin—A degenerated oligonucleotide was designed against the conserved region around the histidine active site residue of subtilisin-like serine proteases, with the sequence GTICCIGCIACRTGIGTICCRTG-3′ (I, inosine; R, A, or G) (primer 1, Fig. 2). Additionally, the NH_2-terminal amino acid sequence of the purified pyrolysin was used to design a degenerated oligonucleotide against residues 5–11 with the sequence, ACITGGGTIATWAAY-GCIYT-3′ (I, inosine; W, A or T; Y, C or T) (primer 2, Fig. 2). PCR reactions were performed using 100 ng of both oligonucleotides and 250 ng *P. furiosus* chromosomal DNA as template (19) in a final volume of 100 ml. After denaturation of the template (5 min 95 °C), 30 cycles (1 min 95 °C, 2 min 35 °C, and 3 min 72°) were performed, followed by a final extension of 7 min at 72 °C on a DNA Thermal Cycler (Perkin-Elmer). Southern hybridization of *P. furiosus* chromosomal DNA was performed with the obtained 0.6-kb PCR fragment. Hybridizing fragments were isolated from agarose gels using GeneClean (BIO 101), cloned into the vector pUC19, and transformed to *E. coli* TG1 by established methods (26).

For the construction of a lambda library, pyrococcal DNA was digested with *Sau*3AI under conditions optimized for the generation of fragments with a size of 15–20 kb that were isolated using Qiaex (Qiagen Inc.), cloned in lambda EMBL3 *Bam*HI-arms provided by Promega and introduced in *E. coli* KW251 following *in vitro* packaging using the Packagene® Lambda DNA Packaging System (Promega Corp.).

RNA Isolation and Primer Extension—Total RNA was isolated from *P. furiosus* cells grown on peptone using guanidinium isothiocyanate as described previously (27). The purified RNA was used as template in primer extension experiments as described previously (19) with oligonucleotide, GTTGGTTAGGGAGTATAGAAGTTG-3′ (primer 3, Fig. 2).

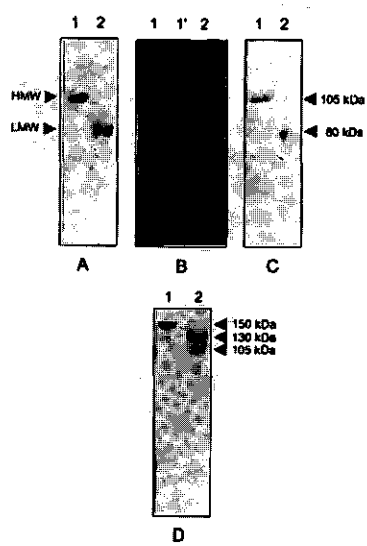


FIG. 1. **Urea-SDS-polyacrylamide gel electrophoresis of purified HMW and LMW pyrolysin.** Purified HMW (*lane 1*) and LMW (*lane 2*) pyrolysin were subjected directly to urea-SDS-PAGE (*A*, *B*, and *C*) or after acid denaturation (*D*). *A* and *D*, proteins stained with Coomassie Brilliant Blue. *B*, activity staining using a Kodak film overlay; *lane 1′* represents HMW pyrolysin after 1-h incubation at 95 °C. *C*, detection of glycosylated proteins using periodic acid-Schiff staining.

Sequence Analysis—Nucleotide sequence analysis of PCR products was carried out by the AmpliTaq® cycle sequencing kit (Perkin-Elmer) and analyzed on a automated DNA sequencer (Applied Biosystems 373A). A Prism™ Ready Reaction DyeDeoxy™ Terminator cycle sequencing kit (Applied Biosystems) was used to sequence both strands of plasmid DNA.

Computer analysis of nucleotide and deduced amino acid sequences were carried out with the PC/GENE program version 5.01 (IntelliGenetics Inc.) and the GCG package version 7.0 (28) at the CAOS/CAMM Centre of the University of Nijmegen (The Netherlands).

Nucleotide Sequence Accession Number—The nucleotide sequence reported has been submitted to the GenBank/EMBL Data Bank with the accession number U55835.

RESULTS

Purification of Pyrolysin—Cells of *P. furiosus* grown on peptides are known to contain high levels of pyrolysin, the highly thermostable serine protease that is mainly associated with the cell envelope (12). The membrane fraction of pyrococcal cells grown on peptone was enriched for pyrolysin via a preincubation in 6 M urea at 95 °C, during which the proteolytic activity remained constant while the total protein concentration decreased approximately 100-fold, most likely due to the proteolytic activity of pyrolysin. The remaining proteins were separated by preparative urea-PAGE followed by a final purification step using anion-exchange chromatography. This resulted in HMW and an LMW pyrolysin fraction (Fig. 1*A*) with apparent molecular masses on urea-SDS-PAGE of 105 and 80 kDa, respectively. Activity staining showed that HMW and LMW pyrolysin each contain one proteolytically active band (Fig. 1*B*).

Characterization of HMW and LMW Pyrolysin—The NH_2-terminal amino acid sequences of both HMW (35 residues) and LMW (5 residues) pyrolysin were determined and found to be

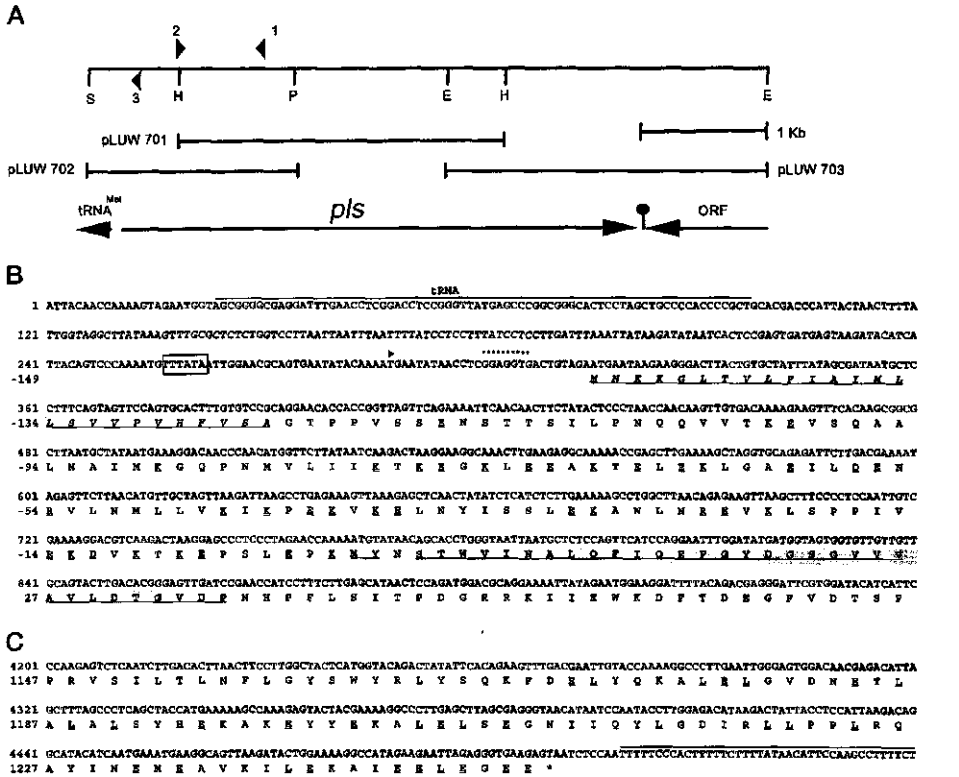


FIG. 2. **Genetic organization and partial nucleotide and amino acid sequence of the *pls* gene.** *A*, physical and genetic map of the *pls* gene and flanking regions with subclones and relevant restriction sites (*S*, *Ssp*I; *H*, *Hin*dIII; *P*, *Pst*I; *E*, *Eco*RI). The *arrowheads* indicate the location and orientation of the primers that were used in PCR (primers 1 and 2) or primer extension experiments (primer 3). *B*, nucleotide and deduced amino acid sequence of the 3′ end of the *pls* gene and its upstream region, including the $tRNA^{Met}$ gene (*overlined*). TATA box of the *pls* gene is indicated; its transcription initiation site is marked by the *arrow*, and its ribosome binding site is shown by the *asterisks*. The signal sequence of the pyrolysin precursor is shown in *italics* and *underlined*, whereas the identified NH_2-terminal amino acid residues of the mature pyrolysin are *underlined* and *shaded*. Charged residues present in the propeptide are *underlined*. *C*, nucleotide and deduced amino acid sequence of the 5′ end of the *pls* gene with the proposed termination sequence composed of stretches of thymidine residues (*overlined*). The predominant Leu and Glu residues are *underlined*.

identical (Fig. 2). The relatedness of both pyrolysin fractions was supported by the observation that the third residue of their NH_2 termini could not be determined following Edman degradation, even after repeated trials (data not shown). Inspection of the amino acid sequence deduced from the *pls* gene (Fig. 2, see below) revealed that the third position of the mature pyrolysin is an asparagine that could be post-translationally modified, since it is the first residue of the well-known *N*-glycosylation motif Asn-*X*-(Ser/Thr) (29). Therefore, periodic acid-Schiff staining was performed on HMW and LMW pyrolysin separated by urea-SDS-PAGE (Fig. 1*C*). The results indicate that both proteins are glycosylated to a similar extent and that LMW is not a deglycosylated form of HMW pyrolysin.

A mutual relation between the HMW and LMW pyrolysin could further be demonstrated by immunological cross-reaction using antisera raised against both purified pyrolysin fractions (data not shown). Moreover, a prolonged incubation of purified HMW pyrolysin at 95 °C resulted in its decrease and the formation of LMW pyrolysin. This is illustrated by the semi-quantitative *in situ* detection of a new proteolytic activity with the size of LMW that originated from HMW pyrolysin, indicating that LMW pyrolysin is a processing product of HMW pyrolysin (Fig. 1*B*). Finally, a multimer structure of HMW pyrolysin was further excluded by applying rigorous acid denaturation conditions followed by separation by urea-SDS-PAGE (Fig. 1*D*). HMW pyrolysin showed a single band with apparent molecular mass of 150 kDa. Two proteins were obtained after acid denaturation of the LMW pyrolysin fraction with apparent molecular masses of 130 and 105 kDa, the smallest of which is most likely a result of autoproteolytic processing of the 130-kDa protein.

Cloning and Characterization of the pls Gene Encoding Pyrolysin—A data base search revealed that the NH_2-terminal amino acid sequence of pyrolysin showed significant similarity with subtilisin-like serine proteases (17). This includes a potential active site aspartate residue at position 30 (Fig. 2, see below), suggesting that pyrolysin is a member of this family of serine proteases. Subtilisin-like serine proteases are characterized by the sequential order of the active site residues Asp, His, and Ser and the high degree of homology on amino acid level around these residues that form the catalytic triad (17). Two degenerated oligonucleotides based on the conserved region surrounding the active site histidine residue of subtilisin-like proteases and on the identified NH_2-terminal amino acid sequence (primers 1 and 2, Fig. 2) were used in a PCR reaction on chromosomal DNA from *P. furiosus*. Sequence analysis of the resulting 0.6-kb PCR fragment (data not shown) indicated that indeed a fragment of a subtilisin-like protease gene was ampli-

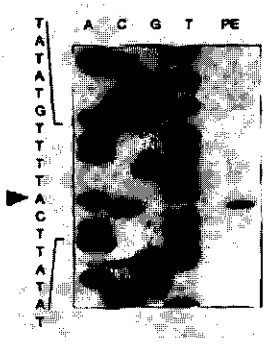


FIG. 3. **Transcription initiation site of the *pls* gene.** Primer extension was performed using primer 3 (*PE*, primer extension product). A DNA sequence ladder was obtained with the same primer and is shown for size comparison (*lanes A, G, C,* and *T*). The *arrow* indicates the deduced transcriptional initiation in the complementary strand.

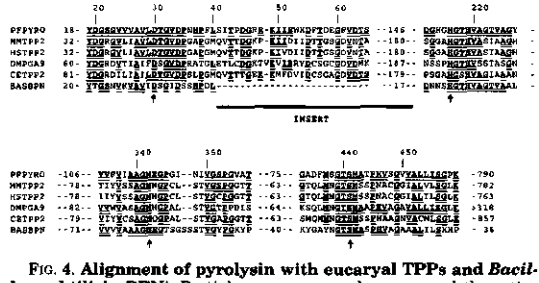


FIG. 4. **Alignment of pyrolysin with eucaryal TPPs and *Bacillus* subtilisin BPN′.** Partial sequences are shown around the active site residues aspartate, histidine, and serine, and the asparagine present in the oxyanion hole (indicated by *arrows*). The number of amino acid residues are given between the *boxes* that are present in the proteins. Large insert is indicated by *solid bar*; *bold and double underlined residues* are conserved between pyrolysin and the other sequences. Sequences were deduced from the following accession numbers: BASBPN, subtilisin BPN′ from *Bacillus amyloliquefaciens* (K02496); MMTPP2, TPP from *Mus musculus* (mouse) (X81323); HSTPP2, TPP from *Homo sapiens* (human) (M73047); DMPGA9, TPP from *D. melanogaster* (R. Nusse and F. van Leeuwen, personal communication); CETPP2, TPP from *C. elegans* (U23176); PFPYRO, pyrolysin from *P. furiosus*.

fied. The 0.6-kb PCR fragment was subsequently used as probe on a Southern blot containing chromosomal DNA of *P. furiosus* digested with various restriction enzymes. A strongly hybridizing 2.7-kb *Hin*dIII-*Hin*dIII fragment was cloned into *Hin*dIII-digested pUC19, resulting in pLUW701 (Fig. 2). Sequence analysis of the insert of pLUW701 revealed that it contained an incomplete ORF, coding for the NH_2-terminal amino acids determined from the mature pyrolysin, however, without a potential translation start codon (Fig. 2). Southern hybridization of *P. furiosus* DNA with a pLUW701-derived probe showed a hybridizing 1.7-kb *Ssp*I-*Pst*I fragment, which was cloned in *Sma*I-*Pst*I-digested pUC19, resulting in pLUW702 (Fig. 2). Sequence analysis of the insert of pLUW702 showed indeed the presence of the 5′ end of the *pls* gene (Fig. 2). The 3′ end of the *pls* gene was cloned from a lambda EMBL3 library consisting of 15–20-kb *Sau*3A fragments of *P. furiosus* DNA. A pool of approximately 2400 independent plaques were screened by hybridization with the *Hin*dIII-*Hin*dIII insert of pLUW701 resulting in 49 hybridizing phages. A 2.5-kb *Eco*RI fragment of one of these phages carrying the 3′ end of the *pls* gene was identified by hybridization and cloned into *Eco*RI-digested pUC19, resulting in pLUW703 (Fig. 2).

Sequence Analysis of the pls Gene, Flanking Regions, and the pls Gene Product Pyrolysin—Sequence analysis of the *pls* gene, contained in pLUW701, pLUW702, and pLUW703, revealed an ORF of 4194 base pairs. The first ATG codon at position 316–319 of this ORF is preceded by a stretch of purine-rich nucleotides (at position 301–308) that may function as a ribosome binding site, suggesting that this is the translation initiation site of the *pls* gene (Fig. 2). If so, the *pls* gene encodes a 1398-residue protein. The NH_2-terminal sequence determined from the purified pyrolysin was found to be encoded by nucleotides 763–868, indicating that the mature NH_2 terminus is preceded by a leader sequence of 149 amino acids (Fig. 2). The first 26 residues of this leader have the characteristics of a signal sequence with a high content of hydrophobic residues that can form a transmembrane-spanning helix and is followed by a putative consensus processing site (30). The remaining 123 residues could code for a propeptide that has a high number (36 residues) of charged residues, predominantly Lys and Glu (17 and 16 residues, respectively) (Fig. 2). The calculated molecular mass of the pyrolysin precursor is 155 kDa and after removal of the leader 138.5 kDa.

Identification of Transcription Initiation Site and Regulatory Sequences—Primer extension experiments were performed on total RNA isolated from pyrococcal cells grown on peptone to identify the transcription initiation site and resulted in a prominent extension product (Fig. 3). The size of this product corresponded with transcription initiation at the thymidine residue located 28 nucleotides upstream of the assumed translation start site (Fig. 2). A hexanucleotide with the sequence TTTATA is located 26 nucleotides upstream of this transcription initiation site and resembles the TATA box, which is involved in binding of the archaeal RNA polymerase (Fig. 2) (31, 32).

Downstream of the *pls* gene a conserved structure is found, composed of stretches of at least four subsequent thymidine residues, that could be involved in termination of transcription (31), which is further followed by another ORF with an unknown function (Fig. 2). Upstream and in opposite orientation of the *pls* gene, the gene for a $tRNA^{Met}$ is located that could be the initiator tRNA, since it contains all the characteristics for an archaeal initiator tRNA (33).

Substrate Specificity—Comparison of the deduced amino acid sequence of pyrolysin with those of other subtilisin-like serine proteases present in the data bases revealed the highest degree of homology (28–32% identity in the catalytic domain) with the tripeptidyl peptidases II (TPPs) from eucaryal origin, *i.e.* human, mouse, *Drosophila melanogaster*, and the nematode *Caenorhabditis elegans* (Fig. 4). The eucaryal TPPs are intracellular proteases with exopeptidase activity (34, 35), whereas all other known subtilisin-like serine proteases are endopeptidases (17).

The substrate specificity of pyrolysin was analyzed with chromogenic peptide substrates (Table I). Both tripeptidyl and tetrapeptidyl substrates were cleaved by pyrolysin. A preference for the conversion of substrates with a positively charged (Lys, Arg) residue at the P1 site was observed (nomenclature according to Schechter and Berger) (36). In spite of its homology with TPPs, pyrolysin showed no activity with the known substrate for TPPs, Ala-Ala-Phe-*p*-nitroanilide, and is not inhibited by the dehydroalanine-containing peptide Arg-Ala-δAla-Val-Ala that has been shown to be a competitive inhibitor of TPPs from human and rat (23, 24). No difference in substrate specificity was found between HMW or LMW pyrolysin (data not shown).

To demonstrate that pyrolysin has endopeptidase activity, it was incubated with α_{S1}-, β- or k-caseins and casein fragments. The degradation of caseins was followed in time and revealed that for all caseins the first cleavage products constituted a

TABLE I
Substrate specificity of pyrolysin

Purified pyrolysin was tested for the degradation of synthetic substrates. <Glu, pyroglutanyl; Suc, succinyl; MeO, methoxy; *p*NA, *p* nitroanilide.

Substrate	Activity
P4 P3 P2 P1[a]	%[b]
Suc-Ala-Ala-Pro-Lys-*p*NA	100
Suc-Ala-Ala-Pro-Arg-*p*NA	32
Suc-Ala-Ala-Pro-Met-*p*NA	13
Suc-Ala-Ala-Pro-Phe-*p*NA	14
Suc-Ala-Glu-Pro-Phe-*p*NA	7
Suc-Ala-Lys-Pro-Phe-*p*NA	65
H-D-Val-Leu-Lys-*p*NA	41
MeO-Suc-Arg-Pro-Tyr-*p*NA	52
<Glu-Pro-Val-*p*NA	0.5
Ala-Ala-Phe-*p*NA	<0.01

[a] Nomenclature according to Schechter and Berger (36).
[b] 100% activity represents a specific activity of 300 nmol/min/mg protein.

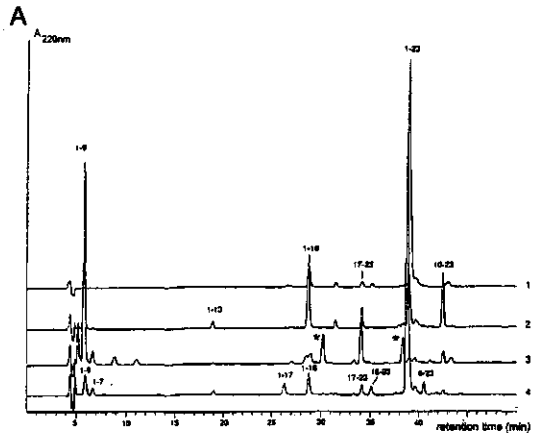


FIG. 5. **Reversed-phase HPLC pattern of degradation products of α_{S1}-casein-(1–23) by pyrolysin.** *A*, pattern of cleavage products resulting from the incubation of α_{S1}-casein-(1–23) with pyrolysin for 0 min (*pattern 1*), 20 min (*pattern 2*), and 40 min (*pattern 3*) at 95 °C. The known pattern (22, 25) of α_{S1}-casein degradation obtained with the P_{III}-type protease from *L. lactis* is shown for comparison (*pattern 4*). Identified products are indicated; *asterisks* indicate new cleavage products after prolonged incubation. *B*, amino acid sequence of α_{S1}-casein-(1–23) with the preferential cleavage sites indicated by the *arrows*.

number of distinct protein bands that were subsequently completely degraded (data not shown). To identify its preferential endopeptidase cleavage sites, pyrolysin was incubated with the peptide α_{S1}-casein-(1–23), and degradation products were analyzed by HPLC (Fig. 5). The results indicate that α_{S1}-casein-(1–23) can be completely degraded by pyrolysin resulting in the appearance of four new peptides that were identified based on their retention time. Prolonged incubation showed a decrease of the peptides 1–16 and 10–23 and the appearance of new cleavage products (indicated as *asterisks* in Fig. 5). The results demonstrate that pyrolysin is an endopeptidase with two preferential cleavage sites in α_{S1}-casein, at bonds 9–10 and 16–17 (Fig. 5).

DISCUSSION

The hyperthermostable serine protease pyrolysin was purified from peptone-grown cells of *P. furiosus* and characterized, and its *pls* gene was cloned via reversed genetics. An essential step in the purification method of pyrolysin from the membrane fraction appeared to be its preincubation in 6 M urea at 95 °C. This resulted in two proteolytic active fractions, HMW and LMW pyrolysin, that were separated by preparative urea-PAGE and shown to have identical NH_2 termini. Incubation of the purified HMW pyrolysin at 95 °C resulted in the formation of LMW pyrolysin, and both forms appeared to be glycosylated to a similar extent. Therefore, the differences in molecular masses of the HMW and LMW pyrolysin are likely to be a consequence of autoproteolytic removal of a COOH-terminal part of the HMW pyrolysin that results in LMW pyrolysin. Similar processing has also been suggested for the serine protease from the related archaeon *Thermococcus stetteri* (11). Moreover, COOH-terminal processing is a common feature of serine proteases (37).

It has frequently been observed that proteins from hyperthermophiles are not denatured during standard SDS-PAGE (3, 38, 39). This was also found for pyrolysin that was subject to autoproteolysis under these conditions. Acid denaturation of HMW pyrolysin demonstrated a molecular mass of 150 kDa. Autoproteolytic processing of pyrolysin may also explain the previously reported heterogeneity of pyrolysin, evidenced by the multiple activity bands on substrate-PAGE (12–14). Alternatively, pyrolysin may form a complex either with itself, as found for TPPs (40), or with a cell envelope structure as found in *Staphylothermus marinus* where a serine protease is associated with a filamentous surface protein assembly called tetrabrachion (39).

The *pls* gene-encoding pyrolysin is most likely transcribed as a single unit, since the transcription initiation site and TATA box could be identified upstream of the gene, whereas downstream it was followed by conserved archaeal termination sequences. Furthermore, upstream of the *pls* gene a $tRNA^{Met}$ gene is found in the opposite orientation that most likely is the initiation tRNA.

Analysis of the deduced amino acid sequence indicates that pyrolysin is synthesized as a prepro-enzyme with a common signal sequence (30), suggesting that export of proteins in *P. furiosus* occurs in a similar way as in Eucarya and Bacteria. The propeptide of pyrolysin shows little homology with the consensus identified for different propeptides of subtilisin-like serine proteases; however, it shares with these propeptides a high number of charged residues (41).

The mature part of the deduced pyrolysin sequence shows the highest homology with eucaryal TPPs, which form a distinct subgroup of the subtilisin-like serine proteases. The mature pyrolysin, as several other subtilisin-like serine proteases, can be divided into two domains, the NH_2-terminal catalytic domain that extends to about residue 500 followed by a COOH-terminal extension to residue 1249 (17). The catalytic domain is characterized by the three active site residues, Asp-30, His-216, and Ser-441, and the presence of the conserved Asn-342 present in the oxyanion hole that stabilizes the reaction intermediate. The catalytic domains of pyrolysin and the TPPs contain a large insert of more than 150 residues between the aspartate and histidine in comparison with the *Bacillus* subtilisin (Fig. 4). A number of conserved residues among all members could be identified within the insert, which may imply an evolutionary relationship between the archaeal pyrolysin and the eucaryal TPPs (Fig. 4). In addition, pyrolysin shows weak homology in the large COOH-terminal domain where the TPPs also have many conserved residues (data not shown).

The COOH-terminal domain, if present, in subtilisin-like serine proteases can have different functions but is often related to localization or anchoring of the protease (17). The last 80 amino acids of pyrolysin do not contain known hydrophobic anchoring sequences, but they are characterized by a high number of Glu and Leu residues (14 and 13 residues, respectively) that may be involved in the observed association of pyrolysin to the cell envelope.

The COOH-terminal domain of pyrolysin together with the large insert present in the catalytic domain contain almost all (29 of 32) of the possible *N*-glycosylation sites (Asn-*X*-(Ser/Thr)) present in pyrolysin. The presence of a gap at position 3 in the determined NH_2-terminal amino acid sequence suggests that Asn at that position is modified, presumably due to glycosylation; whether the other possible glycosylation sites are indeed post-translationally modified and what the structure is of the bound glycan is presently unknown. This form of post-translational modification may function in the thermostabilization of the enzyme as has also been suggested for the extracellular protease from *S. acidocaldarius* (7). Mature pyrolysin without the 16.5-kDa prepropeptide has a calculated molecular mass of 138.5 kDa. The higher apparent molecular mass observed for the acid-denatured HMW pyrolysin (150 kDa) could be explained by glycosylation.

The substrate specificity of pyrolysin was analyzed using chromogenic substrates, different caseins, and the small peptide α_{S1}-casein-(1–23). Degradation of the caseins and α_{S1}-casein-(1–23) into distinct intermediates followed by the complete degradation of these proteins implies that pyrolysin is not an exopeptidase like the eucaryal TPPs but a true endopeptidase like all other subtilisin-like serine proteases. The specificity studies indicated that pyrolysin has a rather broad substrate specificity with a preference for Arg or Lys residues at the P1 position. Such a preference is to be expected when pyrolysin autocatalytically removes its propeptide as is commonly found with serine proteases (42–44) (see Fig. 2). Modelling of pyrolysin on known three-dimensional structures in combination with functional expression and mutational analysis could give further insight in this processing and the molecular mechanisms of adaptation of this extracellular serine protease to extremely high temperatures.

Acknowledgments—We are grateful to Dr. Brigitte Tomkinson for a gift of chemicals and Saskia van Schalkwijk and Marcel Dijkgraaf for technical assistance.

REFERENCES

1. Adams, M. W. W. (1993) *Annu. Rev. Microbiol.* **47,** 627–658
2. Stetter, K. O., Fiala, G., Huber, G., Huber, R., and Segerer, A. (1990) *FEMS Microbiol. Rev.* **75,** 117–124
3. Mukund, S., and Adams, M. W. W. (1995) *J. Biol. Chem.* **270,** 8389–8392
4. Kengen, S. W. M., de Bok, F. A. M., van Loo, N.-D., Dijkema, C., Stams, A. J. M., and de Vos, W. M. (1994) *J. Biol. Chem.* **269,** 17537–17541
5. Kelly, R. M., and Adams, M. W. W. (1994) *Antonie Leeuwenhoek* **66,** 247–270
6. Woese, C. R., Kandler, O., and Wheelis, M. L. (1990) *Proc. Natl. Acad. Sci. U. S. A.* **87,** 4576–4579
7. Fusek, M., Lin, X., and Tang, J. (1990) *J. Biol. Chem.* **265,** 1496–1501
8. Lin, X., and Tang, J. (1990) *J. Biol. Chem.* **265,** 1490–1495
9. Morikawa, M., Izawa, Y., Rashid, N., Hoaki, T., and Imanaka, T. (1994) *Appl. Environ. Microbiol.* **60,** 4559–4566
10. Cowan, D. A., Smolenski, K. A., Daniel, R. M., and Morgan, H. W. (1987) *Biochem. J.* **247,** 121–133
11. Klingeberg, M., Galunsky, B., Sjoholm, C., Kasche, V., and Antranikian, G. (1995) *Appl. Environ. Microbiol.* **61,** 3098–3104
12. Eggen, R., Geerling, A., Watts, J., and de Vos, W. M. (1990) *FEMS Microbiol. Lett.* **71,** 17–20
13. Blumentals, I. I., Robinson, A. S., and Kelly, R. M. (1990) *Appl. Environ. Microbiol.* **56,** 1992–1998
14. Connaris, H., Cowan, D. A., and Sharp, R. J. (1991) *J. Gen. Microbiol.* **137,** 1193–1199
15. Snowden, L. J., Blumentals, I. I., and Kelly, R. M. (1992) *Appl. Environ. Microbiol.* **58,** 1134–1141
16. Volkl, P., Markiewicz, P., Stetter, K. O., and Miller, J. H. (1995) *Protein Sci.* **3,** 1329–1340
17. Siezen, R. J., de Vos, W. M., Leunissen, J. A. M., and Dijkstra, B. W. (1991) *Protein Eng.* **4,** 719–737
18. Gibson, T. J. (1984) *Studies on the Epstein-Barr-Virus Genome.* Ph.D. Thesis. Cambridge, UK
19. Eggen, R. I. L., Geerling, A. C. M., Jetten, M. S. M., and de Vos, W. M. (1991) *J. Biol. Chem.* **266,** 6883–6887
20. Zacharius, R. M., Zell, T. E., Morrison, J. H., and Woodlock, J. J. (1969) *Anal. Biochem.* **30,** 148–152
21. Laemmli, U. K. (1970) *Nature* **227,** 680–685
22. Siezen, R. J., Bruinenberg, P. G., Vos, P., van Alen-Boerrigter, I. J., Nijhuis, M., Alting, A. C., Exterkate, F. A., and de Vos, W. M. (1993) *Protein Eng.* **6,** 927–937
23. Tomkinson, B. (1994) *Biochem. J.* **304,** 517–523
24. Tomkinson, B., Grehn, L., Fransson, B., and Zetterqvist, O. (1994) *Arch. Biochem. Biophys.* **314,** 276–279
25. Exterkate, F. A., Alting, A. C., and Slangen, C. J. (1991) *Biochem. J.* **273,** 135–139
26. Sambrook, J., Fritsch, E. F., and Maniatis, T. (1989) *Molecular Cloning: A Laboratory Manual,* 2nd Ed., Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY
27. Voorhorst, W. G. B., Eggen, R. I. L., Luesink, E. J., and de Vos, W. M. (1995) *J. Bacteriol.* **177,** 7105–7111
28. Devereux, J., Haeberli, P., and Smithies, O. (1990) *Nucleic Acids Res.* **12,** 387–395
29. Marshall, R. D. (1972) *Annu. Rev. Biochem.* **41,** 673–702
30. Von Heijne, G. (1985) *J. Mol. Biol.* **184,** 99–105
31. Dalgaard, J. Z., and Garrett, R. A. (1993) *The Biochemistry of Archaea (Archaebacteria)* (Kates, M., Kushner, D. J., and Matheson, A. T., eds) pp. 535–563, Elsevier, Amsterdam
32. Rowlands, T., Baumann, P., and Jackson, S. P. (1994) *Science* **264,** 1326–1329
33. Kuchino, Y., Ihara, M., Yabusaki, Y., and Nishimura, S. (1982) *Nature* **298,** 684–685
34. Bålöw, R.-M., Tomkinson, B., Ragnarsson, U., and Zetterqvist, Ö. (1986) *J. Biol. Chem.* **261,** 2409–2417
35. Tomkinson, B., Wernstedt, C., Hellman, U., and Zetterqvist, O. (1987) *Proc. Natl. Acad. Sci. U. S. A.* **84,** 7508–7512
36. Schechter, I., and Berger, A. (1967) *Biochem. Biophys. Res. Commun.* **27,** 157–162
37. Shinde, U., Li, Y., Chatterjee, S., and Inouye, M. (1993) *Proc. Natl. Acad. Sci. U. S. A.* **90,** 6924–6928
38. Kengen, S. W. M., and Stams, A. J. M. (1994) *Biocatalysis* **11,** 79–88
39. Peters, J., Nitsch, M., Kuhlmorgen, B., Golbik, R., Lupas, A., Kellermann, J., Engelhart, H., Pfunder, J. P., Muller, S., Goldie, K., and Engel, A. (1995) *J. Mol. Biol.* **245,** 385–401
40. Macpherson, E., Tomkinson, B., Bålöw, R., Höglund, S., and Zetterqvist, Ö. (1987) *Biochem. J.* **248,** 259–263
41. Siezen, R. J., Leunissen, J. A. M., and Shinde, U. (1995) *Intramolecular Chaperones and Protein Folding* (Inouye, M., and Shinde, U., eds) pp. 233–256, R. G. Landes Co, Austin, TX
42. Germain, D., Dumas, F., Vernet, T., Bourbonnais, Y., Thomas, D. Y., and Boileau, G. (1992) *FEBS Lett.* **299,** 283–286
43. Leduc, R., Molloy, S. S., Thorne, B. A., and Thomas, G. (1992) *J. Biol. Chem.* **267,** 14304–14308
44. Power, S. D., Adams, R. M., and Wells, J. A. (1986) *Proc. Natl. Acad. Sci. U. S. A.* **83,** 3096–3100

Chapter 7

Homology Modelling of Two Subtilisin-like Serine Proteases from Hyperthermophilic Archaea *Pyrococcus furiosus* and *Thermococcus stetteri*

Wilfried G.B. Voorhorst, Angela Warner, Willem M. de Vos, and Roland, J. Siezen

reprinted with permission from **Protein Engineering** 1997, vol. 10, 905-914

Protein Engineering vol.10 no.8 pp.905–914, 1997

Homology modelling of two subtilisin-like serine proteases from the hyperthermophilic archaea *Pyrococcus furiosus* and *Thermococcus stetteri*

Wilfried G.B.Voorhorst[1], Angela Warner[1], Willem M.de Vos[1,2] and Roland J.Siezen[2,3]

[1]Department of Microbiology, Wageningen Agricultural University, Hesselink van Suchtelenweg 4, NL-6703 CT Wageningen, The Netherlands and [2]Department of Biophysical Chemistry, NIZO, PO Box 20, NL-6710 BA Ede, The Netherlands

[3]To whom correspondence should be addressed

The hyperthermophilic archaeon *Pyrococcus furiosus* produces an extracellular, glycosylated hyperthermostable subtilisin-like serine protease, termed pyrolysin (Voorhorst,W.G.B., Eggen,R.I.L., Geerling,A.C.M., Platteeuw,C., Siezen,R.J. and de Vos,W.M. (1996) *J. Biol. Chem.*, 271, 20426–20431). Based on the pyrolysin coding sequence, a pyrolysin-like gene fragment was cloned and characterized from the extreme thermophilic archaeon *Thermococcus stetteri*. Like pyrolysin, the deduced sequence of this serine protease, designated stetterlysin, contains a catalytic domain with high homology with other subtilases, allowing homology modelling starting from known crystal structures. Comparison of the predicted three-dimensional models of the catalytic domain of stetterlysin and pyrolysin with the crystal structure of subtilases from mesophilic and thermophilic origin, i.e. subtilisin BPN′ and thermitase, and the homology model of subtilisin S41 from psychrophilic origin, led to the identification of features that could be related to protein stabilization. Higher thermostability was found to be correlated with an increased number of residues involved in pairs and networks of charge–charge and aromatic–aromatic interactions. These highly thermostable proteases have several extra surface loops and inserts with a relatively high frequency of aromatic residues and Asn residues. The latter are often present in putative N-glycosylation sites. Results from modelling of known substrates in the substrate-binding region support the broad substrate range and the autocatalytic activation previously suggested for pyrolysin.

Keywords: Archaea/*Pyrococcus furiosus*/serine protease/*Thermococcus stetteri*/thermostability

Introduction

To gain insight in the molecular basis of thermostability of proteins, structurally homologous proteins with different thermostability have been compared using known crystal structures or predicted structures from homology modelling (Teplyakov *et al.*, 1990; Davail *et al.*, 1994; Heringa *et al.*, 1995; Tomschy *et al.*, 1994; Muir *et al.*, 1995; Szilágyi and Závodszky, 1995). The recent isolation and characterization of proteins and their genes from hyperthermophiles, microorganisms able to grow optimally at 80°C or higher (Stetter *et al.*, 1990), generates the possibility to extend the structural analysis to extremely stable proteins to gain insight in the stabilization mechanisms that allow proteins to be functional near the maximum temperature of life. Analysis of three-dimensional structures and modelled structures of proteins from hyperthermophiles and their comparison with those from less stable proteins has resulted in identification of features that may be related to protein stabilization. Mechanisms by which proteins from thermophiles are stabilized resemble those identified for mesophilic model proteins and include hydrogen bonds, hydrophobic interactions, ion-pairs, secondary structure stabilization, cavity filling, reduction of the length of loops and conformational strain (reviewed by Tomschy *et al.*, 1994; Völkl *et al.*, 1994; Muir *et al.*, 1995; Szilágyi and Závodsky, 1995; Korndorfer *et al.*, 1995; Yip *et al.*, 1995; Vieille and Zeikus, 1996; Knapp *et al.*, 1997). It seems that each thermostable protein is stabilized by a unique combination of different mechanisms resulting in a more rigid structure. A common theme that has now emerged from analysis of the structures identified so far for hyperthermophiles suggests that an increased number of ion-pairs, that may be arranged in extensive networks, contributes to the observed extreme thermostability (Korndorfer *et al.*, 1995; Yip *et al.*, 1995; Knapp *et al.*, 1997). Moreover, many proteins from hyperthermophiles are not only extremely thermostable, but also show a higher resistance to chemical denaturation and proteolytic degradation than their less stable counterparts (Fontana, 1991). This can be illustrated with pyrolysin, an extracellular serine protease from the archaeon *Pyrococcus furiosus* that shows optimal growth at 100°C (Fiala and Stetter, 1986). This protease has a half-life value of 4 h at 100°C and is still active in the presence of 6 M urea (Eggen *et al.*, 1990; Voorhorst *et al.*, 1996).

The family of subtilisin-like serine proteases, also referred to as subtilases (Siezen *et al.*, 1991; Siezen and Leunissen, 1997) has been extensively studied from psychrophilic, mesophilic and thermophilic origin to gain insight in thermostability, catalytic activity and substrate specificity (Wells and Estell, 1988; Davail *et al.*, 1994; Daniel *et al.*, 1995; Feller *et al.*, 1996). However, knowledge of subtilases from hyperthermophilic origin is limited to pyrolysin from *P.furiosus* (Eggen *et al.*, 1990; Voorhorst *et al.*, 1996) and STABLE from *Staphylothermus marinus* (Mayr *et al.*, 1996), which have been extensively characterized on the biochemical level with specific attention for thermostability and substrate specificity. Moreover, the subtilase aerolysin from the hyperthermophile *Pyrobaculum aerophilum* has been analysed for features that could be related to its thermostability using homology modelling (Völkl *et al.*, 1994).

The amino acid sequence homology in the catalytic domain of pyrolysin with other subtilases is sufficiently high to allow for homology modelling (Voorhorst *et al.*, 1996). However, the identification of features as being important for protein stabilization based on only a single enzyme may lead to misinterpretation (Querol *et al.*, 1996). Therefore, we have

now isolated and characterized the gene encoding a homolog of pyrolysin, designated stetterlysin, from the extreme thermophilic archaeon *Thermococcus stetteri* that grows optimally at 75°C (Miroshnichenko *et al.*, 1989). The amino acid sequence homology within the core of the catalytic domain between stetterlysin, pyrolysin and other subtilases allowed homology modelling using known three-dimensional protein structures of subtilisin BPN′ from the mesophilic bacterium *Bacillus amyloliquefaciens* with an optimum growth temperature between 30 and 40°C (Heinz *et al.*, 1991; Gallagher *et al.*, 1995) and the more thermostable thermitase from *Thermoactinomyces vulgaris* that grows optimally at 50°C (Gros *et al.*, 1989). The predicted three-dimensional structures of the highly thermostable pyrolysin and stetterlysin were compared with the homology model of subtilisin S41 from the psychrophilic *Bacillus* T41 (optimal growth temperature of 4°C), that has a half-life value of only a few minutes at 50°C (Davail *et al.*,1994), and with subtilisin BPN′ and thermitase, with half-life values of 23 min at 60°C (Braxton and Wells, 1992) and 1 h at 85°C (Daniel *et al.*, 1995), respectively. Conformational characteristics of the catalytic domain of stetterlysin and pyrolysin were analysed that could be related to their thermostability regarding overall composition, ion-pair formation, aromatic–aromatic interactions, amino acid replacements, helix stabilization and calcium-binding sites. Additionally, the substrate binding regions of these archaeal subtilases were modelled in detail with respect to substrate specificity and interactions between the enzymes and their known substrates.

Materials and methods

Bacterial strains, plasmids and enzymes

Cells of *Thermococcus stetteri* (DSM 5262) grown on tryptone were kindly provided by Dr G.Antranikian (Klingeberg *et al.*, 1995). *Escherichia coli* TG1 (Gibson, 1984) was cultured in LB-based medium at 37°C. The plasmids pUC18, pUC19 (Pharmacia Biotechnologies) and pGEM^{R}-T (Promega Corporation) were used as cloning vectors. Standard molecular biological techniques were performed as described by Sambrook *et al.* (1989). Restriction enzymes, DNA modifying enzymes and *Taq* DNA polymerase were acquired from Life Technologies Inc. Oligonucleotides were purchased from Pharmacia Biotechnologies.

Isolation and cloning of plsT gene coding for stetterlysin from T.stetteri

Degenerated oligonucleotides were designed against the conserved regions around the Asp and Ser active-site residues of subtilisin-like serine proteases in combination with the known sequence of pyrolysin from *P.furiosus* (Voorhorst *et al.*, 1996) and include BG95(Asp): 5′-GTIGCIGTIMTIGAYACIGG-3′ and BG31(Ser): 5′-GGISYIGCCATISWGTICC-3′ (I, inosine; M, A or C; Y, C or T; S, G or C; W, A or T). PCR reactions were performed using 100 ng *T.stetteri* chromosomal DNA (Eggen *et al.*, 1991) as template in a final volume of 100 µl. After denaturation of the template (5 min, 95°C) amplification was performed for 35 cycles (1 min, 95°C; 2 min, 35°C and 3 min, 72°C), followed by a final extension of 7 min at 72°C on a DNA Thermal Cycler (Perkin Elmer Cetus). This resulted in a prominent product of 1.4 kb that hybridized with a pyrolysin-derived probe. The 1.4 kb PCR fragment was isolated from agarose gel using Gene Clean (BIO 101), cloned into the vector pGEM^{R}-T and transformed to *E.coli* TG1. Nucleotide sequence analysis showed the presence of a *Kpn*I restriction site in the amplified fragment that could be used for cloning chromosomal DNA fragments. A Southern blot containing chromosomal DNA of *T.stetteri* digested with different combinations of *Kpn*I and other restriction enzymes was probed with the cloned PCR fragment. Positively hybridizing fragments were subsequently cloned in appropriately digested pUC18 and characterized by nucleotide sequence analysis.

Sequence analysis and alignment

Automated nucleotide sequence analysis of plasmid DNA was carried out on either an Applied Biosystems 373A sequencer (Applied Biosystems Inc.) using a Prism^{TM} Ready Reaction DyeDeoxy^{TM} Terminator cycle sequencing kit or the Li-Cor 4000L sequencer (Li-Cor Inc.) using the Amersham Thermo Sequenase cycle sequencing kit with 7-deaza-dGTP and infrared labelled oligonucleotides (MWG-Biotech, Germany). Computer analysis of nucleotide and deduced amino acid sequences was carried out with the PC/GENE program version 5.01 (IntelliGenetics Inc.) and the GCG package version 7.0 (Devereux *et al.*, 1990) at the CAOS/CAMM Centre of the University of Nijmegen (The Netherlands).

The amino acid sequences of pyrolysin and stetterlysin were included in the multiple sequence alignment of the catalytic domains of the entire subtilase family (Siezen *et al.*, 1991; Siezen and Leunissen, 1997). In this way the sites of insertions in the catalytic domain could be identified. The numbering of subtilisin BPN′ indicated by stars (*) was used as reference.

Modelling of the catalytic domains of pyrolysin and stetterlysin

Homology modelling was performed using QUANTA 3.2.3 (Molecular Simulations, Cambridge, UK) (Sali and Blundell, 1993) and CHARMm 22 (Brooks *et al.*, 1983) running on a Silicon Graphics 4D25TG workstation. The model is based on sequence homology between pyrolysin, stetterlysin and members of the subtilase family as reflected in the given alignment (Figure 1), and on the known X-ray structures of subtilisin BPN′ in complex with the inhibitor R45-eglin (Heinz *et al.*, 1991), thermitase in complex with eglin (Gros *et al.*, 1989) and subtilisin in complex with its pro-peptide (Gallagher *et al.*, 1995). Insertions of more than six residues relative to either subtilisin or thermitase were not modelled. After addition of H-atoms the entire molecule was subjected to energy minimization after constraining the active site residues (Asp32*, His64* and Ser221*), the Asn155* of the oxy-anion hole, and the two β-sheet strand backbones eI and eIII involved in substrate binding. The atomic coordinates of the two models are electronically available (ftp://ftp.caos.kun.nl/pub/molbio/siezen97/).

The backbone conformation of eglin residues 40–47 or subtilisin pro-peptide residues –6 to –1 (i.e. the segment that interacts with subtilisin or thermitase in the crystal structures) was used to model enzyme–substrate interactions of pyrolysin by substituting the side chains by the appropriate P6-P2′ residues—nomenclature as described (Schechter and Berger, 1967)—of either pro-pyrolysin or of α_{S1}-casein residues 11–18.

Nucleotide accession numbers

The nucleotide sequence reported has been submitted to the GenBank/EMBL Data Bank with the accession number AF015225.

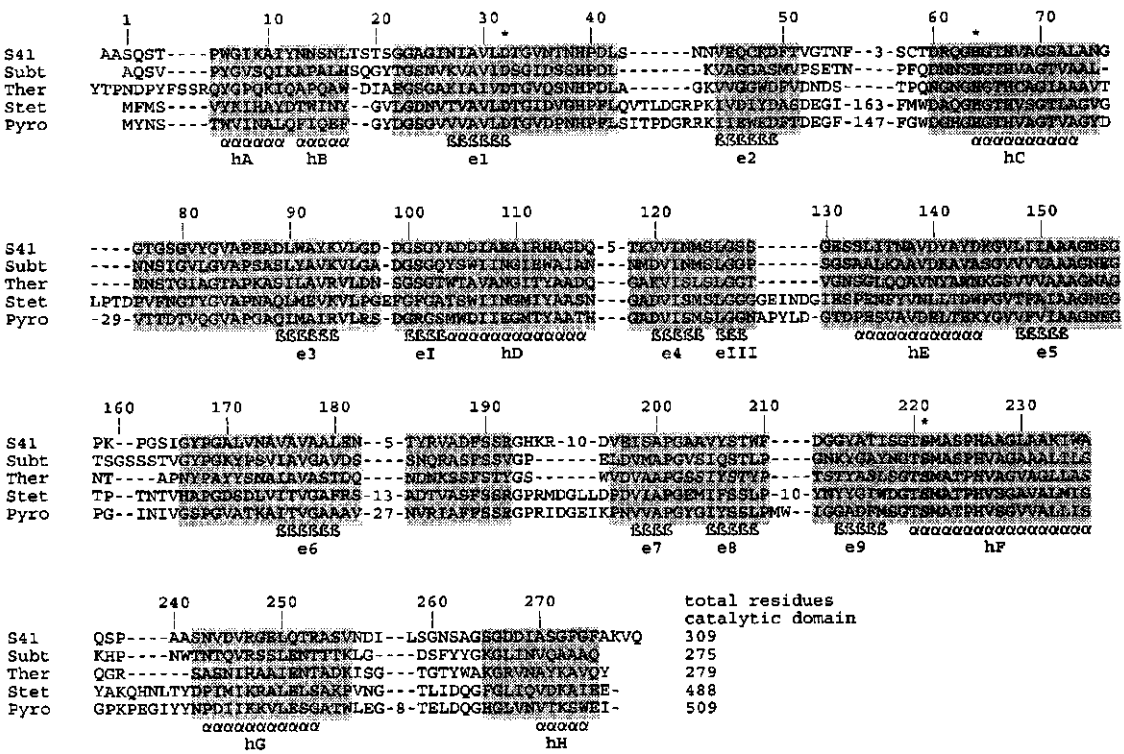


Fig. 1. Amino acid sequence alignment of the catalytic domains of subtilisin BPN′ (Subt), thermitase (Ther), subtilisin S41 (S41), stetterlysin (Stet) and pyrolysin (Pyro). The number of not-modelled residues in inserts is indicated. Catalytic residues D32*, H64* and S221* are indicated by (*). Structurally conserved core regions of subtilisin/thermitase are shaded, and common secondary structure elements are labelled (Siezen *et al.*, 1991; Siezen and Leunissen, 1997). The numbering of subtilisin BPN′ is used.

Results

Isolation and characterization of the pyrolysin-like gene from T.stetteri

To expand our knowledge of highly stable subtilases, the extreme thermophilic archaeon *T.stetteri* was screened for the presence of a serine protease similar to pyrolysin from *P.furiosus* (Voorhorst *et al.*, 1996). A PCR approach was used with degenerate oligonucleotides derived from conserved amino acids surrounding the active site Asp and Ser of subtilases and pyrolysin. From the obtained PCR products a 1.4 kb fragment was the most prominent one and hybridized with a pyrolysin-derived probe. This PCR fragment was cloned and sequence analysis indicated that indeed a fragment coding for a pyrolysin-like serine protease was amplified. The cloned PCR fragment was used as hybridization probe to identify *T.stetteri* DNA fragments that encoded the pyrolysin-like serine protease. Two adjacent fragments, a 2.9 kb *Sph*I–*Kpn*I and a 1.5 kb *Kpn*I–*Bam*HI fragment, were cloned in appropriately digested pUC18. Nucleotide sequence analysis of these cloned *T.stetteri* fragments confirmed the sequence of the PCR fragment and revealed the presence of a 2667 bp open reading frame, starting with an ATG and preceded by a purine rich region that may constitute the ribosome binding site; no stop codon was found to terminate this open reading frame. Its deduced sequence of 888 amino acids contained a complete catalytic domain of a serine protease, that we have designated stetterlysin after its source *T.stetteri*.

Homology between stetterlysin and pyrolysin

The sequence of stetterlysin showed the highest overall homology with a subgroup of the subtilase family that includes the archaeal pyrolysin and tripeptidyl-peptidases from eucaryal origin (Voorhorst *et al.*, 1996; Siezen and Leunissen, 1997). The predicted organization of stetterlysin also showed high similarity to that of pyrolysin and starts with a prepro-peptide of 159 residues (26% identity to pyrolysin) followed by a 488 residue N-terminal catalytic domain (46% identity) and an incomplete C-terminal extension of at least 241 residues (30% identity). The catalytic domain of these archaeal proteases, as other subtilases, can be divided into the *core*, residues that are also present in subtilisin and/or thermitase, and the *inserts*, residues that are additional to the core (Siezen *et al.*, 1991). The core of the archaeal proteases stetterlysin and pyrolysin showed high amino acid sequence homology (up to 56% similarity, and 44% identity) with other subtilases (Figure 1). Both pyrolysin and stetterlysin have an unusually large insert, almost one-third of the catalytic domain residues, just preceding the catalytic His. These inserts show 39% identity, but are not homologous to any known proteins or specifically to the large inserts at the equivalent position in tripeptidyl peptidases,

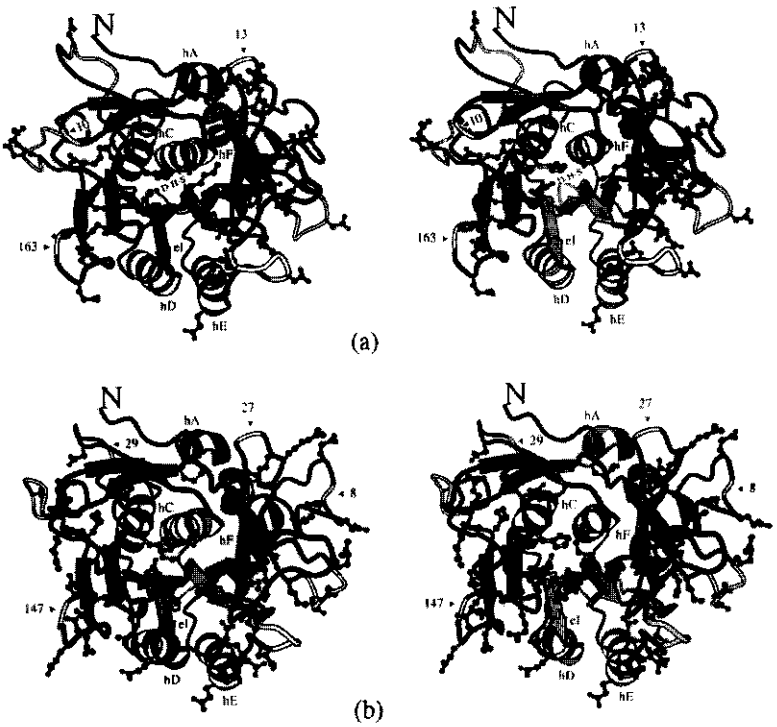


Fig. 2. Stereo plot of three-dimensional models of stetterlysin (**a**) and pyrolysin (**b**) catalytic domain core showing ionic interactions. Only the negatively and positively charged side chains are shown, coloured red and blue, respectively. The core structure backbone is shown in green and positions of inserts in yellow; the number of not-modelled residues in these inserts is indicated. Catalytic residues Asp, His and Ser are shown in the centre, and the substrate binding region is between the two β-strands (light blue) in the front. Drawn with MOLSCRIPT (Kraulis, 1991).

suggesting that stetterlysin and pyrolysin are most closely related in an evolutionary sense (Siezen and Leunissen, 1997).

Homology models

Three-dimensional structure models for stetterlysin and pyrolysin were constructed based on the crystal structures of subtilases from mesophilic and thermophilic origin, respectively subtilisin BPN' and thermitase, and on the sequence alignment of the core residues of stetterlysin and pyrolysin with subtilisin and thermitase (Figure 1). Inserts in stetterlysin and pyrolysin of more than six residues, and in particular the large inserts of 163 residues in stetterlysin and 147 residues in pyrolysin between the catalytic Asp and His residues, could not be modelled. The homology modelling resulted in predicted three-dimensional models of the catalytic domain core (Figures 2 and 3) for which over 92% of the residues (excluding Gly, Pro) had phi-psi angles in most favoured or allowed regions in a Ramachandran plot; moreover, side chain parameters were all within acceptable limits as assessed by PROCHECK (Laskowski *et al.*, 1993; data not shown). The alignment and the predicted three-dimensional models were used to analyse amino acid composition (Table I) and structural features of the catalytic domain that could be related to thermostability, as discussed below.

Charged residues and ion pair formation

The frequency of charged residues in the core of the subtilases increased from subtilisin via thermitase to the two archaeal proteases stetterlysin and pyrolysin, with a doubling of negatively charged residues (Table I). Moreover, the inserts in these archaeal subtilases contain an even higher percentage (14%) of negatively charged residues, specifically caused by Asp residues. As a consequence, these archaeal proteases have an overall excess of negatively charged residues in the catalytic domain, which is also reflected in its low (calculated) pI (3.8–4.0), both with and without the inserts (Table I).

The predicted three-dimensional structures of the catalytic core of stetterlysin and pyrolysin were analysed for possible ion-pair formation (Figure 2), defined as oppositely charged amino acid side chains in the vicinity of each other, preferably within a distance of 4.0 Ångstrom (Barlow and Thornton, 1983). The number of charged residues involved in such ion pairs increased from subtilisin (9 residues), thermitase (16 residues), stetterlysin (21 residues) to pyrolysin (34 residues) (Table II) (see also Siezen *et al.*,1991). In stetterlysin and pyrolysin a number of these charged residues are predicted to be arranged in such a way that one charged residue has multiple interactions with oppositely charged residues resulting

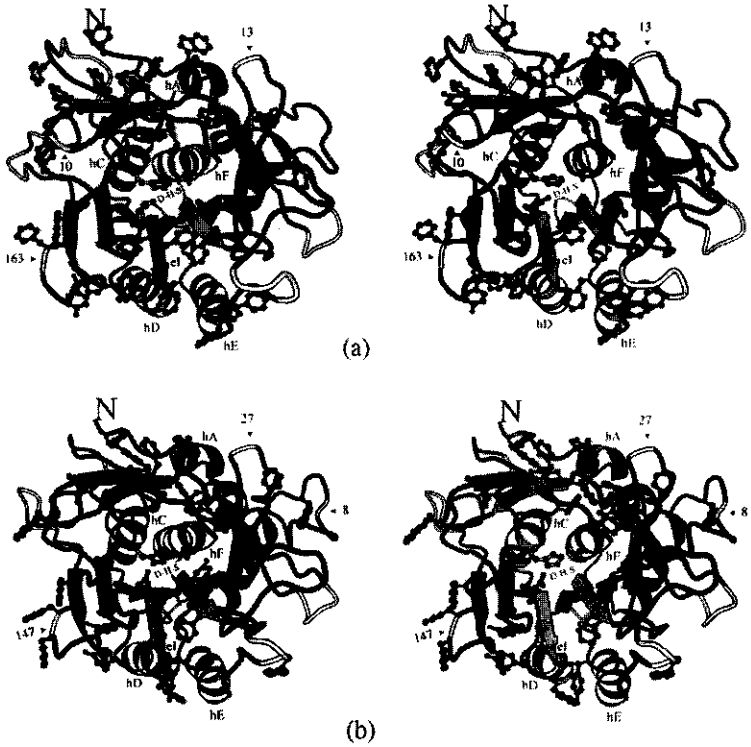


Fig. 3. Stereo plot of three-dimensional models of stetterlysin (**a**) and pyrolysin (**b**) catalytic domain showing aromatic–aromatic interactions. Only the aromatic side chains are shown, in purple. Other details as in Figure 2.

in ion-pair networks. These include for instance K247* in stetterlysin and E143* in pyrolysin (Table II). The predicted ion-pair networks were found to cover the surface of the catalytic core of both enzymes (Figure 2). Only two ion pairs, R247*–D197* and R247*–E251*, are found to be highly conserved. The ion pair R247*–D197* is conserved among subtilisin, thermitase and stetterlysin and subtilisin S41 from psychrophilic origin (Davail *et al.*, 1994), while the pair R247*–E251* is conserved among subtilisin, thermitase, stetterlysin and pyrolysin. Another ionic interaction is conserved among the thermostable thermitase, stetterlysin and pyrolysin, and includes the residues D49*, R94* and D52*; these residues contribute to Ca^{2+} ion binding in thermitase (Gros *et al.*, 1989).

Aromatic residues and interactions

The frequency of aromatic residues in subtilases increased from mesophilic (5.8%) to hyperthermophilic origin (13.8%) (Table I). Moreover, the inserts in the stetterlysin and pyrolysin were found to be extremely rich in aromatic residues, particularly those of pyrolysin that contained 20% aromatic residues. In contrast to what is often found in globular proteins, these aromatic residues were all more or less exposed to the outside of the domain core (Figure 3). The only exception to this was the internal Phe (F149*) in stetterlysin that could be involved in improvement of packing efficiency.

The moderately conserved aromatic residues for all subtilases at position 6, 50, 113 and 189 (Siezen *et al.*, 1991) were also found to be conserved in pyrolysin and stetterlysin, except for the equivalent of residue 50 in stetterlysin (Figure 1).

The side chain of an aromatic residue can interact with the side chain of another aromatic residue by weak non-bonded forces, leading to so-called aromatic–aromatic interactions as described by Burley and Petsko (1985). The surrounding of the side chain of each aromatic residue in the predicted structure of stetterlysin and pyrolysin was analysed for the possibility to be involved in such aromatic–aromatic interactions (Figure 3). Since nearly all of these aromatic residues were located at the surface of the domain core, they hardly encounter any conformational constraints, and hence it is likely that aromatic residues in the vicinity of each other show an interaction in the native protein. The number of aromatic residues predicted to be involved in such aromatic–aromatic interactions is 7, 14, 14 and 15 for respectively subtilisin, thermitase, stetterlysin and pyrolysin (Table II) (see also Siezen *et al.*, 1991). No aromatic pairs are highly conserved and each enzyme appears to use a different set of aromatic residues for stabilization.

Amino acid composition and replacements

The frequency of polar residues in the core of stetterlysin and pyrolysin was found to be decreased by almost a third compared with the less thermostable subtilases, subtilisin and thermitase (Table I). The frequency of Ala and Ser residues decreased about twofold going towards higher thermostability, whereas

Table I. Amino acid frequencies (%)[a] in the catalytic domain.

			Stetterlysin			Pyrolysin		
Amino acids	Subtilisin	Thermitase	(core)	(insert)	(total)	(core)	(insert)	(total)
Isoelectric point	275 pI = 6.6	279 pI = 7.4	273 pI = 4.1	215 pI = 3.6	488 pI = 3.8	273 pI = 4.7	236 pI = 3.5	509 pI = 4.0
Charged	12.4	12.1	16.2	18.1	17.0	18.3	17.0	17.7
Acidic	5.5	5.4	10.3	14.0	11.9	10.3	13.6	11.8
Basic	4.7	5.4	3.7	2.8	3.3	5.5	3.0	4.3
Polar	28.4	31.8	19.4	24.2	21.5	19.4	22.1	20.6
Hydrophobic	36.7	35.1	39.2	30.1	35.5	38.1	33.1	35.8
Aromatic	5.8	8.7	9.6	13.9	12.5	8.5	19.9	13.8
A Ala	13.5	15.8	9.5	3.3	6.8	8.8	4.2	6.7
C Cys	0.0	0.4	0.0	0.0	0.0	0.0	0.0	0.0
D Asp	3.6	4.7	7.0	10.2	8.4	5.9	8.9	7.3
E Glu	1.8	0.7	3.3	3.7	3.5	4.4	4.7	4.5
F Phe	1.1	1.1	4.8	3.7	4.3	3.7	5.1	4.3
G Gly	12.0	11.8	12.1	9.8	11.1	14.3	8.9	11.8
H His	2.2	1.4	2.2	1.4	1.8	2.6	0.4	1.6
I Ile	4.7	5.0	8.1	5.6	7.0	7.3	5.1	6.3
K Lys	4.0	3.6	2.6	0.5	1.6	3.3	1.3	2.4
L Leu	5.5	3.2	5.9	9.3	7.4	4.8	7.2	5.9
M Met	1.8	0.4	3.7	3.3	3.5	2.6	1.3	2.0
N Asn	6.6	7.5	4.4	8.4	6.2	3.7	6.8	5.1
P Pro	5.1	4.3	5.5	4.2	4.9	4.7	5.1	4.9
Q Gln	3.6	4.3	1.8	3.3	2.5	1.8	2.1	2.0
R Arg	0.7	1.8	1.1	2.3	1.6	2.2	1.7	2.0
S Ser	13.5	10.8	7.0	2.8	5.1	7.0	5.9	6.5
T Thr	4.7	7.9	6.2	9.8	7.8	7.0	7.2	7.1
V Val	10.9	7.9	9.2	7.4	8.4	11.4	9.3	10.4
W Trp	1.1	2.2	1.8	1.4	1.6	2.2	2.1	2.2
Y Tyr	3.6	5.4	4.0	9.8	6.6	2.6	12.7	7.3

[a] Charged, RKHDE; acidic, DE; basic, KR; polar, NCQST; hydrophobic, AILFWV; aromatic: FWY.

their frequency in the inserts decreased even further (Table I). Many of the Ala residues present in thermitase were specifically replaced by Leu or Val in stetterlysin and pyrolysin (data not shown).

Specific replacement of residues such as Lys→Arg, Gly→Ala, Ser→Thr and Asp→Glu have been suggested to be related to thermostability (Argos *et al.*, 1979; Menéndez-Arias and Argos, 1989). The modelled proteases were analysed for these specific replacements, but none of these was found to occur to any significant extent (data not shown). Moreover, for basic residues only substitutions in the opposite direction, i.e. Arg→Lys, were observed at three positions from thermitase to stetterlysin. Nevertheless, the overall replacement ratio R/(R+K) increased from subtilisin (0.15) to pyrolysin (0.4), suggesting a preference for Arg residues that are more suitable to form multiple stabilizing ion pairs.

Helix stabilization

Compensation of α-helix dipoles can lead to stabilization of a protein (Querol *et al.*, 1996). Analysis of the helices in stetterlysin and pyrolysin showed that except for helix D (nomenclature Siezen *et al.*, 1991; Siezen and Leunissen, 1997), all helices appeared to have a negatively charged residue, mainly Asp, near the N-terminus of the helix that may function as N-cap to compensate the helix dipole. At the C-terminus of the helices no preference in amino acid residues could be identified. The compensation of the dipole at the N-terminus has also been observed for the hyperthermostable aerolysin (Völkl *et al.*, 1994). However, the subtilase S41 from psychrophilic origin also showed N-helix capping (Davail *et al.*, 1994). Therefore, the role of N-helix capping in protein thermostabilization is doubtful (Völkl *et al.*, 1994).

Changing the helix composition was found to be another feature contributing to protein stabilization (Nicholson *et al.*, 1988; Fersht and Serrano, 1993). Ala residues have been found to be prevalent in α-helices of thermostable enzymes and were also identified as stabilizing this secondary structure (Menéndez-Arias and Argos, 1989; Fersht and Serrano, 1993). In the archaeal proteases stetterlysin and pyrolysin no correlation was observed between the number of Ala residues in the helices and thermostability, comparable with what was observed for the citrate synthase from *P.furiosus* (Muir *et al.*, 1995). This contrasts with the situation described for the hyperthermostable aerolysin from *P.aerophilum* where transitions to Ala at the ends of α-helices were suggested to be related to thermostability (Völkl *et al.*, 1994). However, for stetterlysin and pyrolysin some of the other described stabilizing interactions, ion pairs and aromatic–aromatic interaction were found within some of the helices (Figure 2 and 3).

Ca^{2+} ion binding

The binding of calcium ions in a protein can lead to stabilization by reducing the local flexibility. For the family of subtilases a number of calcium binding sites are known: three of these, Ca1, Ca2 and Ca3, are occupied by a Ca^{2+} ion in thermitase (Gros *et al.*, 1989; Siezen *et al.*, 1991). None of these calcium binding sites was found to be present as such in stetterlysin or pyrolysin, since suitable ligands are not present at the equivalent positions in the sequence. However, a few ligands at the Ca2 site were found to be conserved, viz D49* and D52*; these may be involved in calcium binding in combination

Table II. Ionic and aromatic interactions deduced from crystal structures of subtilisin BPN′ (Subt) and thermitase (Ther), and predicted from homology modelling of stetterlysin (Stet) and pyrolysin (Pyro)

	Residues	Subt	Ther	Stet	Pyro		Residues	Subt	Ther	Stet	Pyro
Ionic:	10–184		K–D			Aromatic:					
	42[+7]–42[+5]			R–D	R–D		4–206		Y–Y		
	42[+8]–46				R–E		4–214		Y–Y		
	48–109				K–E		4[+1]–17		F–W		
	87–22		K–E				6–206				W–Y
	94–49		R–D	K–D	R–D		6–214				W–F
	94–52		R–D	K–D	R–D		21–274				Y–W
	97–106				R–D		47–113				W–Y
	101–99				R–D		48–50		W–F		
	136–140	K–D					48–113		W–Y	Y–Y	
	141–112	K–E					50–113		F–Y		F–Y
	144–143				K–E		57–59			F–W	F–W
	145–116		K–D				91–113	Y–W			
	170–195	K–E					99–101			F–F	
	173–140				K–E		101–106			F–W	
	173–143				K–E		136–137			F–Y	
	181–203			R–E			137–144			Y–W	
	185–258				R–E		167–170		Y–Y		
	192–262			R–D	R–D		167–171	Y–Y	Y–Y		
	194[+1]–197			R–D			188–189				F–F
	194[+1]–251			R–E			188–203				F–Y
	194[+1]–260				R–E		189–203				F–Y
	195–194[+3]				K–D		192–262		Y–Y		
	239–239[+2]				K–E		206–214			F–Y	
	239–275[+1]			K–E			211–213			Y–Y	
	247–194[+5]				K–E		213–214			Y–Y	
	247–197	R–D	R–D	K–D			240–241				Y–Y
	247–251	R–E	R–E	K–E	R–E		261–262	F–Y			
	248–244				K–D		262–263	Y–Y	Y–W		
	248–275[+1]			R–E							
	255–12			K–D							
	267–184		R–D								
	267–255		R–D								
	272–255		K–D								
	272–271			K–D							
	272–275			K–E	K–E						

Numbering of subtilisin BPN′ is used as reference throughout; numbers between brackets refer to the inserts (see also Siezen *et al.*, 1991).

with ligands originating from the inserts that are closely located to this site. Moreover, other calcium binding sites in either stetterlysin or pyrolysin may well be present within inserts, since the enzyme activity of pyrolysin is known to be affected by EDTA (Eggen *et al.*, 1990).

Substrate binding and specificity of pyrolysin

For pyrolysin the preferential endopeptidase cleavage sites have been determined using the small peptide α_{S1}-casein(1–23) and showed a fairly broad specificity with preference for cleavage after large residues, both polar and non-polar (Voorhorst *et al.*, 1996). The substrate binding region was also predicted by homology modelling of the catalytic domain of pyrolysin (Figure 4a). Pyrolysin may activate itself autocatalytically by removal of its own propeptide, as has been demonstrated for other members of the subtilisin-like serine protease family (Power *et al.*, 1986; Germain *et al.*, 1992; Leduc *et al.*, 1992). Therefore, the P4–P2′ segment around the pro-peptide cleavage site in precursor pyrolysin was modelled as substrate as well as the previously identified cleavage site at bond 16–17 in α_{S1}-casein(1–23) (Voorhorst *et al.*, 1996) (Figure 4a). The interaction with the pro-peptide is also shown in detail in a stereo view (Figure 4b).

In subtilisin and thermitase the substrate specificity appears to be largely determined by interactions of the substrate P4 and P1 residue side chains with the large, hydrophobic S4 and S1 binding subsites (Gron *et al.*, 1992). In pyrolysin these two subsites are predicted to be similar to subtilisin, also suggesting a preference for large hydrophobic P1 and P4 side chains such as Phe, Tyr, Leu and Met. In addition, long side chains with a polar end group may also be preferred since they could form H-bonds with N129* and S167* at the bottom of the S1 and S4 subsites, respectively. Electrostatic interactions between enzyme and substrate could occur in the S1 subsite where the E156* should lead to a preference for basic P1 residues (Lys, Arg), and at the S3 subsite where the R101* should lead to a preference for acidic P3 residues (Glu, Asp). The previously observed substrate specificity with chromogenic substrates is generally in good agreement with this model, although positively charged P3 residues appear to be even more favourable than negatively charged ones (Voorhorst *et al.*, 1996).

The two larger peptide substrates modelled meet the P1–P4 requirements very well (Figure 4). The α_{S1}-casein fragment has large hydrophobic (Leu) and polar residues (Gln) at the P1 and P4 positions and an acidic residue (Glu) at P3. In pro-pyrolysin the P4–P2′ residues around the putative cleavage site, i.e. Leu-Glu-Pro-Lys↓Met-Tyr, appear to form very favourable hydrophobic and charge interactions within the substrate-binding region (Figure 4), providing support for autocatalytic activation of pyrolysin. An increased number of electrostatic interactions between pyrolysin and its substrate may explain

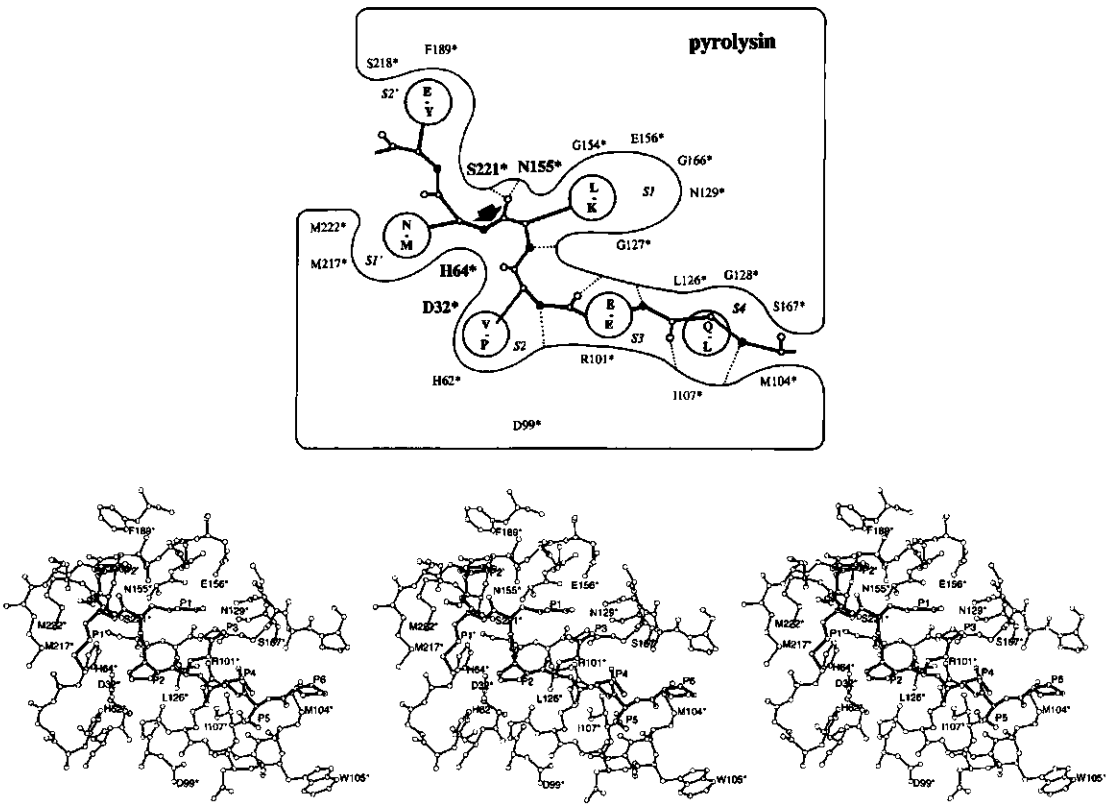


Fig. 4. (a) Schematic representation of the substrate binding region of pyrolysin. Catalytic residues are D32*, H64* and S221*, and N155* is the oxyanion hole residue (large bold). Binding subsite nomenclature S4–S2′ according to Schechter and Berger (1967). The two substrates depicted are α_{s1}-casein(13–18) (Gln-Glu-Val-Leu↓Asn-Glu), at the top, and pro-pyrolysin residues −4 to +2 (Leu-Glu-Pro-Lys↓Met-Tyr) at the bottom; cleavage site after the P1 residue is indicated by a large arrow. Substrate backbone atoms are shown as solid spheres (nitrogen), small open spheres (carbon) and large open spheres (oxygen); potential backbone hydrogen bonds between the substrate and pyrolysin are dotted. (b) Stereo view of the modelled substrate binding region of pyrolysin (thin lines) in complex with pro-pyrolysin residues P6–P2′ (thick lines).

the observed higher thermostability of pyrolysin in the presence of substrate (Figure 4b) (data not shown). It is a general feature of enzymes from mesophiles and thermophiles that the presence of substrate confers enhanced stability to the protein.

Substrate binding modelling of stetterlysin

The substrate binding region predicted for stetterlysin appeared to be somewhat different from that of pyrolysin and had the substitutions H62*Q, D99*E, R101*F, M104*T, G166*H and M217*W (compare to Figure 4a). The main difference was the presence of a H166* residue at the end of the S1 subsite, suggesting that the enzyme may have preference for cleavage after a negatively charged residue. Cytolysin from *Enterococcus faecalis* is the only other subtilase with a His residue at the same position (Siezen and Leunissen, 1997) and it has a very specific substrate specificity with cleavage after a Glu residue (Booth *et al.*, 1996).

The predicted substrate binding region and the postulated autoproteolytic cleavage site (after Lys-Leu-Pro-Asp) of stetterlysin suggest a substrate specificity with preference for a negatively charged residue at the P1 position to generate a stabilizing interaction between the substrate and the enzyme.

Discussion

Proteins from hyperthermophiles

For the identification of features important in protein thermostabilization comparative structural data are required for related proteins with differences in thermostability. Numerous conformational characteristics related to thermostability have been identified in this way (reviewed by Daniel *et al.*, 1995; Goldman, 1995; Ó Fágáin, 1995; Rees and Adams, 1995; Daniel, 1996; Querol *et al.*, 1996; Vieille and Zeikus, 1996). Combined with the hypothesis that the first organisms on this planet were hyperthermophiles (Stetter, 1994), and that therefore the organisms grown at lower temperature are in fact cold adapted, it is essential to include proteins from hyperthermophilic organisms in structural comparisons to identify features related to protein stabilization. To date no subtilases from hyperthermophiles have been crystallized. Therefore, three-dimensional models of the catalytic domain core were predicted by homology modelling using known crystal structures. The homology modelling was initiated with pyrolysin, but for the identification of conserved features related to protein thermostabilization the modelling approach

Table III. Features of the catalytic domain of subtilases that may be related to thermostability

	S41	Subtilisin	Thermitase	Stetterlysin	Pyrolysin
Opt. growth temperature host (°C)	4	30	50	75	100
Half-life values	3 min/50°C	23 min/60°C	1 h/85°C	–	4 h/100°C
Charge-charge interactions[a] (# residues)	4	9	16	21	34
Aromatic-aromatic interactions[a] (# residues)	0	7	14	14	15
Disulfide bonds	1	0	0	0	0
Frequency of Asn + Gln (%)	9.7	10.2	11.8	8.7	7.1
Possible N-glycosylation sites in inserts	–	–	–	7	8

[a]Excluding inserts

was extended with stetterlysin, a pyrolysin homolog from *T.stetteri*. The predicted three-dimensional models of the catalytic domain core of these two highly thermostable subtilases allowed for a detailed analysis of substrate binding and prediction of the specificity. In the case of pyrolysin this confirmed the experimentally determined autocatalytic cleavage site (Voorhorst *et al.*, 1995). Moreover, the deduced models of stetterlysin and pyrolysin complete the temperature range for subtilases that have an identified or predicted three-dimensional structure, allowing identification of features related to thermostabilization of the core of subtilases as summarized in Table III. The temperature range starts at the lower temperature for life with the subtilisin S41 from psychrophilic origin and via subtilisin BPN′ and the more thermostable thermitase to the upper temperature for life with the highly thermostable stetterlysin and hyperthermostable pyrolysin.

Residues sensitive to thermal destruction

Disulphide bridges have been identified to contribute to the overall stability of proteins and the introduction of disulphide bridges can enhance thermal stability (Masazumi *et al.*, 1989). However, proteins from hyperthermophiles in general show a strong decrease in the number of cysteine residues, since at higher temperatures cysteines are sensitive to thermal destruction and disulphide bridges tend to scramble (Gupta, 1991; Vieille and Zeikus, 1996). Therefore, it is not surprising that stetterlysin and pyrolysin have no cysteine residues.

For other amino acid residues sensitive to thermal destruction, such as Asn and Gln, a reduction in abundance is observed in the catalytic domain core of the highly thermostable proteases stetterlysin and pyrolysin as also has been observed for other proteins from hyperthermophiles (Vieille and Zeikus, 1996). However, the inserts in the catalytic domain of these archaeal proteases have a higher frequency of Asn residues, despite the fact that they are often located at the surface of the protein. Closer analysis of the local amino acid sequence shows that the increased frequency is due to Asn residues that are present in the well known N-glycosylation motif, Asn-X-(Ser/Thr) (see below).

Ion pairs

Charged residues contributing to ion pairs have been identified as one of the major factors that can influence protein stability (for review see Querol *et al.*, 1996; Vieille and Zeikus, 1996). A correlation has been found with the increased number of ion pairs and a higher thermostability, but also the formation of ion pair networks and thermostability (Tomschy *et al.*, 1994; Szilágyi and Závodsky, 1995; Yip *et al.*, 1995). For the glutamate dehydrogenase from *P.furiosus* a network which included 18 ion pairs was found, with residues having multiple interactions with oppositely charged residues (Yip *et al.*, 1995). For the proteases compared in our work there is a large increase in the number of residues involved in ion pair and network formation in the core of subtilases from mesophilic to hyperthermophilic origin (Table III). This can be further extended to psychrophiles since subtilisin S41 has even less residues involved in ion pair formation (Davail *et al.*, 1994). Calculation of the number of residues involved in ion pair formation causes an underestimation of the exact number of ion pairs in the native protein, especially since multiple interactions of a charged residue seem to increase towards higher temperatures (Yip *et al.*, 1995). However, the modelling does not allow a further refinement in the exact number of ion–ion interactions.

Simultaneously with this increased number of residues involved in ion–ion interactions towards thermostable enzymes, there is a dramatic decrease down to zero of positively charged residues that are not involved in such interactions. The observed increase of negatively charged residues in the catalytic domain of the studied archaeal subtilases is due to a large increase in Asp residues that are mainly located at the surface of the catalytic domain. This may not be regarded as an adaptation to high temperature, since the subtilase S41 from psychrophilic origin shows a similar frequency of Asp residues (Davail *et al.*, 1994). Modelling studies of a malate dehydrogenase from halophilic origin suggests stabilization by a hydration shell formed by exposed negatively charged residues covering the entire protein surface that may protect against aggregation at high salt concentrations (Dym *et al.*, 1995).

Aromatic interactions

Analysis of the amino acid composition of the different proteases analysed in this study and their (predicted) structures strongly suggests a further positive correlation in thermostability of the enzyme with the presence of aromatic residues that are involved in interactions in the catalytic domain (Table III). Whereas subtilisin S41 from psychrophilic origin has no aromatic–aromatic interactions (Davail *et al.*, 1994), subtilisin has seven and thermitase has 14 residues that are involved in aromatic–aromatic interactions that are important in the enhanced stability of the structure (Teplyakov *et al.*, 1990; Siezen *et al.*, 1991). The predicted core structures of stetterlysin and pyrolysin contain an equal number of aromatic residues at the surface of the core compared with thermitase and most of them are oriented in such a manner that they may be able to interact or even form clusters as found in thermitase (Figure 3) (Teplyakov *et al.*, 1990; Siezen *et al.*, 1991). Additionally, the inserts within the catalytic domain of stetterlysin and pyrolysin, that are in loops on the surface of the core, contain a high percentage of aromatic residues. This suggests that in the total catalytic domain even more aromatic–aromatic interactions may occur to stabilize the structure. That

aromatic–aromatic interactions may be important in protein stabilization has also been deduced from the high percentage of aromatic residues in proteins from *P.woesei*, which has an optimal growth temperature of 100°C (Hensel, 1993). Additionally, the frequency of aromatic residues in thermostable β-glucosidases was also found to be higher than in their mesophilic homologues (Voorhorst *et al.*, unpublished results).

Histidine residues have also been reported to interact with aromatic residues and thereby stabilize the protein (Fersht and Serrano, 1993). For stetterlysin and pyrolysin such interactions seem to be present and possibly extend the size of the aromatic clusters (data not shown).

Stabilization of the external regions

The length of loops connecting elements of regular secondary structure has been suggested to be important for thermostability, such that reducing the length would result in a more compact and therefore more stable enzyme (Russel *et al.*, 1994). In contrast, the hyperthermostable proteases pyrolysin and stetterlysin contain large inserts in the catalytic domain. For the other hyperthermostable archaeal subtilases, STABLE and aerolysin, structural core elements were also found to be connected by longer instead of shorter sequences (Völkl *et al.*, 1994; Mayr *et al.*, 1996). At the same time these loops (or inserts) of all these archaeal subtilases contain a relatively high number of Asn residues, which are known to be sensitive to thermal destruction. It is known that deamidation of Asn residues can be delayed or prevented by appropriate hydrogen bonding to the peptide backbone. On the other hand, various of these Asn residues are part of the possible N-glycosylation site Asn-X-(Ser/Thr), suggesting that these Asn residues may be post-translationally modified (Table III). Indeed, all of the highly stable extracellular archaeal proteases have been found to be glycosylated, including the serine proteases pyrolysin and STABLE (Mayr *et al.*, 1996; Voorhorst *et al.*, 1996). This suggests that such modifications may be an additional way to stabilize the structure and prevent (auto)proteolysis.

Acknowledgements

We thank Jack Leunissen (CAOS/CAMM Centre, Nijmegen, The Netherlands) and Bernadette Renckens for assistance with homology modelling and design of figures. Part of this work was supported by contracts BIOT-CT93-0274 and BIOT-CT96-0488 of the European Union.

References

Argos,P., Rossmann,M.G., Grau,U.M., Zuber,H., Frank,G. and Tratschin,J.D. (1979) *Biochemistry*, **18**, 5698–5703.
Barlow,D.J. and Thornton,J.M. (1983) *J. Mol.Biol.*, **168**, 867–885.
Booth,M.C., Bogie,C.P., Sahl,H., Siezen,R.J., Hatter,K.L. and Gilmore,M.S. (1996) *Mol. Microbiol.*, **21**, 1175–1184.
Braxton,S. and Wells,J.A. (1992) *Biochemistry*, **31**, 7796–7801.
Brooks,B.R., Bruccoleri,R.E., Olafson,B.D., States,D.J., Swaninathan,S. and Karplus,M. (1983) *J. Comp. Chem.*, **4**, 187–217.
Burley,S.K. and Petsko,G.A. (1985) *Science*, **229**, 23–28.
Daniel,R.M., Toogood,H.S. and Bergquist,P.L. (1995) *Biotechn. Gen. Engng. Rev.*, **13**, 51–102.
Daniel,R.M. (1996) *Enzyme Microb. Tech.*, **19**, 74–79.
Davail,S., Feller,G., Narinx,E. and Gerday,C. (1992) *Gene*, **116**, 143–144.
Davail,S., Feller,G., Narinx,E. and Gerday,C. (1994) *J. Biol. Chem.*, **269**, 17448–17453.
Devereux,J., Haeberli,P. and Smithies,O. (1990) *Nucleic Acids Res.*, **12**, 387–395.
Dym,O., Mevarech,M. and Sussman,J.L. (1995) *Science*, **267**, 1344–1346.
Eggen,H.I.L., Geerling,A.C.M., Watts,J. and de Vos,W.M. (1990) *FEMS Microbiol. Lett.*, **71**, 17–20.
Eggen,R.I.L., Geerling,A.C.M., Jetten,M.S.M. and de Vos,W.M. (1991) *J. Biol. Chem.*, **266**, 6883–6887.
Fersht,A.R. and Serrano,L. (1993) *Curr. Opin. Struct. Biol.*, **3**, 75–83.
Feller,G. Narinx,E., Arpigny,J.L., Aittaleb,M., Baise,E., Genicot,S. and Gerday,C. (1996) *FEMS Microbiol. Rev.*, **18**, 189–202.
Fiala,G. and Stetter,K.O. (1986) *Arch. of Microbiol.*, **145**, 56–61.
Fontana,A., (1991) In de Presco,G. (ed.), *Life under Extreme Conditions: Biochemical Adaptation*. Springer-Verlag, Berlin, Germany, pp. 89–113.
Gallagher,T., Bryan,P. and Gilliland,G.L. (1993) *Proteins*, **16**, 205–213.
Gallagher,T., Gilliland,G., Wang,L. and Bryan,P. (1995) *Structure*, **3**, 907–914.
Germain,D., Dumas,F., Vernet,T., Bourbonnais,Y., Thomas,D.Y. and Boileau,G. (1992) *FEBS Lett.*, **299**, 283–286.
Gibson,T.J. (1984) PhD thesis, Cambridge University, Cambridge, UK.
Goldman,A. (1995) *Structure*, **3**, 1277–1279.
Gron,H., Meldal,M. and Breddam,K. (1992) *Biochemistry*, **31**, 6011–6018.
Gros,P., Betzel,C., Dauter,Z., Wilson,K.S. and Hol,W.G.J. (1989) *J. Mol. Biol.*, **210**, 347–367.
Gupta,M.N. (1991) *Biotech. Appl. Biochem.*, **14**, 1–11.
Heinz,D.W., Priestle,J.P., Rahuel,J., Wilson,K.S. and Grütter,M.G. (1991) *J. Mol. Biol.*, **217**, 353–371.
Hensel,R. (1993) In Kates,M., Kushner,D.J. and Matheson,A.T. (eds), *The Biochemistry of Archaea (Archaebacteria)*. Elsevier, NY, pp. 209–211.
Heringa,J., Argos,P., Egmond,R. and de Vlieg,J. (1995) *Protein Engng.*, **8**, 21–30.
Klingeberg,M., Galunsky,B., Sjoholm,C., Kasche,V. and Antranikian,G. (1995) *Appl. Environ. Microbiol.*, **61**, 3098–3104.
Korndorfer,I., Steipe,B., Huber,R., Tomschy,A. and Jaenicke,R. (1995) *J. Mol. Biol.*, **246**, 511–521.
Knapp,S., de Vos,W.M., Rice,D. and Ladenstein,R. (1997) *J. Mol. Biol.* **267**, 916–932.
Kraulis,P.J. (1991) *J. Appl. Crystallogr.*, **24**, 946–950.
Laskowski,R.A., MacArthur,M.W., Moss,D.S. and Thornton,J.M. (1993) *J. Appl. Crystallogr.*, **26**, 283–291.
Leduc,R., Molloy,S.S., Thorne,B.A. and Thomas,G. (1992) *J. Biol. Chem.*, **267**, 14304–14308.
Nicholson,H., Becktel,W.J. and Matthews,B.W. (1988) *Nature*, **336**, 651–656.
Masazumi,M., Signor,G. and Matthews,B.W. (1989) *Nature*, **342**, 291–293.
Mayr,J., Lupas,A., Kellermann,J., Eckerskorn,C., Baumeister,W. and Peters,J. (1996) *Curr. Biol.*, **6**, 739–749.
Menéndez-Arias,L. and Argos,P. (1989) *J. Mol. Biol.*, **206**, 397–406.
Miroshnichenko,M.L., Bonch-Osmolovskaya,E.A., Neuner,A., Kostrikina,N.A., Chernych,N.A. and Alekseev,V.A. (1989) *System. Appl. Microbiol.*, **12**, 257–262.
Muir,J.M., Russell,R.J.M., Hough,D.W. and Danson,M.J. (1995) *Protein Engng.*, **8**, 583–592.
Ó Fágáin,C. (1995) *Biochem. Biophys. Acta*, **1252**, 1–14.
Power,S.D., Adams,R.M. and Wells,J.A. (1986) *Proc. Natl Acad. Sci. USA*, **83**, 3096–3100.
Querol,E., Perez-Pons,J.A. and Mozo-Villarias,A. (1996) *Protein Engng.*, **9**, 265–271.
Rees,D.C. and Adams,M.W.W. (1995) *Structure*, **3**, 251–254.
Russell,R.J.M. Hough,D.W., Danson,M.J. and Taylor,G.L. (1984) *Structure*, **2**, 1157–1167.
Sali,A. and Blundell,T.L. (1993) *Mol. Biol.*, **234**, 779–815.
Sambrook,J. Fritsch,E.F. and Maniatis,T. (1989), *Molecular Cloning: A Laboratory Manual*, 2nd Ed. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.
Schechter,I. and Berger,A. (1967) *Biochem. Biophys. Res. Commun.*, **27**, 157–162.
Siezen,R.J., de Vos,W.M., Leunissen,J.A.M. and Dijkstra,B.W. (1991) *Protein Engng.*, **4**, 719–737.
Siezen,R.J. and Leunissen,J.A.M. (1997) *Protein Sci.*, **3**, 501–523.
Stetter,K.O., Fiala,G., Huber,G., Huber,R. and Segerer,A. (1990) *FEMS Microbiol. Rev.*, **75**, 117–124.
Stetter,K.O. (1994) In Bengtson,S. (ed.), Nobel symposium No. 84, Columbia University Press, NY, pp. 143–151.
Szilágyi,A. and Závodsky,P. (1995) *Protein Engng.*, **8**, 779–789.
Teplyakov,A.V., Kuranova,I.P., Harutyanyan,E.H., Vainshtein,B.K., Frömmel,C., Höhne,W.E. and Wilson,K.S. (1990) *J. Mol. Biol.*, **214**, 261–279.
Tomschy,A., Böhm,G. and Jaenicke,R. (1994) *Protein Engng.*, **7**, 1471–1478.
Vieille,C. and Zeikus,G.G. (1996) *Trends Biotech.*, **14**, 183–190.
Völkl,P., Markiewicz,P., Stetter,K.O. and Miller,J.H. (1994) *Protein Sci.*, **3**, 1329–1340.
Voorhorst,W.G.B., Eggen,R.I.L., Geerling,A.C.M., Platteeuw,C., Siezen,R.J. and de Vos,W.M. (1996) *J. Biol. Chem.*, **271**, 20426–20431.
Wells,J.A. and Estell,D.A. (1988) *Trends Biochem. Sci.*, **13**, 291–297.
Yip,K.S.P., Stillman,T.L., Britton,K.L. Artymiuk,P.J., Baker,P.J., Sedelnikova,S.E. Engel,P.C., Pasquo,A., Chiaraluce,R., Consalvi,V., Scandurra,R. and Rice,D.W. (1995) *Structure*, **3**, 1147–1158.

Received February 4, 1997; revised May 8, 1997; accepted May 14, 1997

Chapter 8

Summary and Concluding Remarks

Hyperthermophiles are recently discovered microorganisms which are able to grow optimally above 85° C. Most hyperthermophiles belong to the *Archaea,* the third domain of life. One of the main interests in hyperthermophiles to deepen the insight in the way their proteins are stabilized and how to apply this knowledge to improve the stability of biotechnologically relevant enzymes. In this thesis attention has been focused on hydrolytic enzymes from hyperthermophilic archaea to provide insight in the way these microorganisms stabilize their proteins and are able to perform catalysis around the normal boiling point of water as well as to functionally produce these enzymes in a mesophilic heterologous host. Additionally, the organization and expression of a number of genes in a hyperthermophilic archaeon were studied. Members of two different classes of hydrolytic enzymes, that represent key enzymes in the metabolism of hyperthermophilic archaea, have been characterized at the molecular level: (i) glycosyl hydrolases that are required for growth on ß-linked sugars, and (ii) serine proteases that are involved in the growth on proteins and peptides. The most extensively studied hyperthermophilic archaeon *Pyrococcus furiosus* was used as model organism for the work described in this thesis.

Chapter 1 gives a brief introduction into different aspects of hyperthermophilic archaea, including an overview of the metabolism of polymeric substrates and the hydrolytic enzymes involved, a listing of the mechanisms by which hydrolases and other enzymes from hyperthermophiles are stabilized, and the main characteristics of their molecular biology. The main part of the thesis deals with the *celB* locus (Fig. 1) of *P.furiosus* and serine proteases of *P.furiosus* and the related *Thermococcus stetteri.*

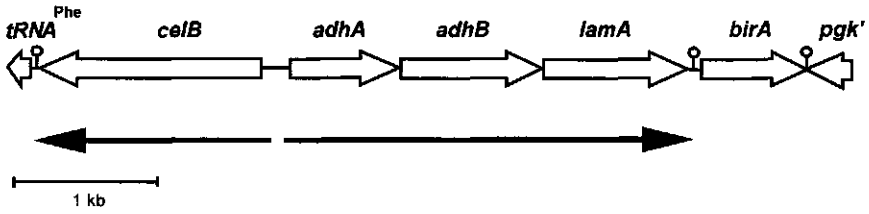


Figure 1. Genetic and transcriptional organization of the *P.furiosus celB* locus.

In chapter 2 the isolation and characterization of the extremely thermostable ß-glucosidase (half-life of 3 days at 100°C) and its *celB* gene of *P.furiosus* is described. The transcriptional organization of the *celB* gene was analyzed and indicated a single transcriptional unit, which was sustained by Northern blot analysis (see below) (Fig. 1). The deduced amino sequence of the ß-glucosidase showed that it is a member of the glycosyl hydrolase family 1. The pyrococcal *celB* gene was overexpressed in the mesophilic host *Escherichia coli* using the *tac* promoter, which resulted in a high production level of ß-glucosidase (up to 20% of the total soluble proteins). The *P.furiosus* ß-glucosidase was produced in an active form by *E.coli* with kinetic and stability characteristics identical to that of the native pyrococcal enzyme. The production of the hyperthermostable ß-glucosidase in a mesophilic host allowed for a simple purification procedure consisting of a heat-treatment of the cell-extracts followed by a single column chromatography. The high production level of the ß-glucosidase in the genetically well-accessible host *E.coli* allowed for protein engineering to gain insight in structure-stability and function relations. Mutational analysis of the active site of the ß-glucosidase from *P.furiosus* showed that the mechanisms for catalysis near the boiling point of water does not differ from that used by homologous enzymes from the family 1 of glycosyl hydrolases optimally active at lower temperatures.

Over the last years, the potential of the extremely thermostable *P.furiosus* ß-glucosidase as a biocatalyst has been studied. Due to its broad substrate specificity and high chemical and thermal stability a wide variety of products could be in glucoconjugation and transglycosylation reactions in water and organic solvents at temperatures in between 75 to 95°C (Fischer *et al.*, 1996; Trincone *et al.*, 1997). The ß-glucosidase from *P.furiosus* has also been analysed for its potential as biocatalyst in the production of ß-galacto-oligosaccharides from lactose (Jansen *et al.*, 1997). Furthermore, the *celB* gene has been developed into a genetic marker to study plant-bacterium interactions and in competition experiments with differently marked *Rhizobium* strains (Sessitsch *et al.*, 1996). The applicability of this system is exemplified by the development of the CelB Gene Marking Kit (FAO/IAEA).

Due to the high production level of the pyrococcal ß-glucosidase in *E.coli* and its simple purification procedure, sufficient pure protein has been generated to initiate crystalization experiments in collaboration with the group of Prof. G.E. Schulz (University of Freiburg) and resulted in crystals that diffract to 3.5 Ångstrom resolution (Schulz, 1997). Efforts are under way to elucidate a high resolution three-dimensional structure of β-glucosidase from *P.furiosus* guided by the recently determined structure

of the homologous *Sulfolobus solfataricus* β-glycosidase LacS in conjunction with a mutational approach to improve the packaging of the pyrococcal crystals and resolve uncertain regions.

Analysis of the genomic region preceding the *celB* gene of *P.furiosus*, revealed the presence of two tandem genes, oppositely orientated of *celB* (Fig. 1). These have been designated *adhA* and *adhB*, since their predicted products showed high homology to short-chain and iron-containing NADP(H)-dependent alcohol dehydrogenases (ADHs), respectively. In chapter 3 these two distinct types of ADHs are studied. The AdhA of *P.furiosus* showed a high degree of conservation in its primary structure with bacterial and eucaryal short-chain ADHs, suggesting that a short-chain ADH was present in the last common ancestor. The AdhB is a member of the group of iron-containing ADHs that is characterized by the presence of four conserved histidine residues. The *adhA* and *adhB* genes were overexpressed in *E.coli* resulting in functional proteins. AdhB showed NADP(H)-dependent activity with methanol as substrate, but this activity was rapidly lost preventing a detailed characterization. The purified AdhA showed a stable NADP(H)-dependent activity towards a broad range of primary and secondary alcohols. AdhA revealed an optimum activity for substrates with a carbon chain of 5 residues, with a preference for the oxidation of secondary over primary alcohols. A higher affinity for aldehydes than for alcohols was observed for AdhA and the pH optimum of this reaction is near the optimum pH for growth of *P.furiosus*. Therefore, it is most likely that the physiological role of AdhA is the reduction of the aldehyde to the corresponding alcohol thereby removing reduction equivalents and regenerating the cofactor. Given the localization of the *adhA* and *adhB* genes, in the vicinity of the *lamA* and *celB* genes, it tempting to speculate that AdhA and AdhB both have a function in the metabolism of ß-linked glucose polymers.

In chapter 4 the molecular characterization of an endo-ß-1,3-glucanase and its *lamA* gene from *P.furiosus* is described. The *lamA* gene was found to be located downstream of the tandem *adhA-adhB* genes within the *celB* locus (Fig. 1). The endo-ß-1,3-glucanase is the first archaeal member of the glycosyl hydrolases family 16 that is composed of endo-ß-1,3 and endo-ß-1,3-1,4-glucanases. The *lamA* gene, without the coding sequence for its N-terminal leader, was cloned behind the T7-promoter in *E.coli*. Using this expression system a functional and extremely stable endo-ß-1,3-glucanase was produced in *E.coli* up to 15 % of the total soluble protein. The purified endoglucanase showed the highest activity on the ß-1,3-glucose polymer laminarin, but has also activity on the ß-1,3-1,4-glucose polymers lichenan and ß-glucan. The

pyrococcal endoglucanase showed optimal activity at pH 6-6.5 and the temperature for maximum activity was 100-105°C, while at 100°C it has a half-life of 19 h. Amino acid sequence alignment of the glycosyl hydrolases of family 16 showed two subgroups; one with endo-ß-1,3-1,4-glucanase activity and another with predominantly endo-ß-1,3-glucanase activity. The latter group is characterized by an additional methionine residue in the predicted active site. Removal of this methionine in the pyrococcal endo-ß-1,3-glucanase by protein engineering did not alter its substrate specificity, but only the catalytic activity.

It was found that *P.furiosus* was able to grown on laminarin and that the endoglucanase *in vitro* hydrolysed laminarin into oligomers and glucose. However, the hydrolysis proceeded more efficiently in combination with the ß-glucosidase of *P.furiosus*. These observations suggest a key role for the extracellular endoglucanase as well as the intracellular ß-glucosidase in the utilization of ß-1,3-linked glucose polymer laminarin.

Chapter 5 describes the regulation of transcription of the divergent *celB* gene and the *adhA-adhB-lamA* gene cluster in *P.furiosus*. Northern blot showed that the *adhA-adhB-lamA* gene cluster form a 2.8-kb operon, designated the *lamA* operon. This operon is flanked upstream by the *celB* gene in opposite orientation and downstream by the *birA* gene, with the *celB* gene being transcribed as monocistronic messengers (Fig. 1). The expression of the enzymes encoded by the *celB* gene and the *lamA* operon of *P.furiosus* was found to be largely dependent on the carbon source present and was highest when the pyrococcal cells were grown on the ß-linked glucose polymers cellobiose and laminarin. The *celB* gene and the divergently orientated *lamA* operon were found to be controlled at the transcriptional level. Moreover, the transcripts were co-regulated and induced by growth on ß-linked glucose polymers. The transcription initiation sites of the *celB* gene and the *lamA* operon were found to be separated by a relatively small intergenic spacer of 142 nucleotides that included the back-to-back promoters, that showed a high degree of conservation, including identical archaeal TATA-box sequences. Transcriptional analysis using an *in vitro* transcription system for *P.furiosus* revealed that both transcripts are initiated from their *in vivo* initiation site. However, the efficiency of transcription initiation was significantly lower than that of the *gdh* gene of *P.furiosus*, suggesting that a positive regulator may contribute to increase the efficiency of transcription of the divergent *celB* gene and *lamA* operon in *P.furiosus*. If so, this would be a bacterial type of regulation. A putative regulatory binding-site could be identified upstream of the promoters of the *celB* gene and *lamA*

operon. These findings open the way use the cell-free transcription system of *P.furiosus* for the identification of the regulator that is likely to be involved in regulation of the divergently orientated genes of the *celB* locus. Alternatively, *in vivo* analysis of the transcriptional control of the *celB* locus may become feasible with the recent development of a transformation system in *P.furiosus* (Aagaard *et al.*, 1996).

A combination of the current knowledge and ideas of the utilization of ß-linked glucose polymers and the transcriptional regulation in the hyperthermophilic archaeon *P.furiosus* led to the following working model (Fig. 2). Large polymeric substrates containing β-1,3-linked glucose residues, such as laminarin are depolymerized to smaller fragments by the extracellular endo-β-1,3-glucanase (LamA). After uptake these fragments can be degraded by the intracellular β-glucosidase activity of CelB. The complementary activity of these two hydrolases has been shown *in vitro*. The co-regulation of transcription of the genes encoding these hydrolases support their concerted action in polymere degradation.

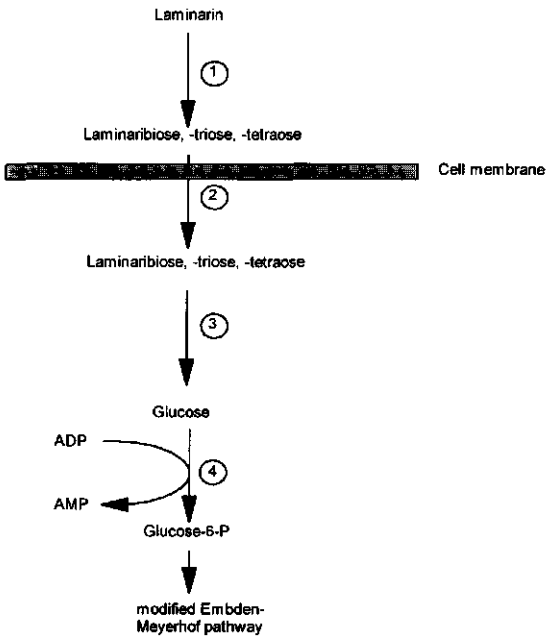


Figure 2. Model for the utilization of β-linked glucose polymers, like laminarin. 1, the extracellular endo-β-1,3-glucanase (LamA); 2, unknown uptake system for oligosaccharides; 3, β-glucosidase (CelB); 4, ADP-dependent hexokinase.

The two remaining experimental chapters deal with the characterization and molecular modelling of proteases involved in the growth on proteinaceous substrates. In chapter 6 the isolation and characterization of the hyperthermostable serine protease, pyrolysin, and its gene from *P.furiosus* are presented. The extracellular pyrolysin was found to be associated to the cell membrane of *P.furiosus*. The major purification step for pyrolysin was obtained via an autoincubation of the membrane fraction in 6 M urea at 95°C, during which a 100-fold purification was obtained. The purification procedure resulted in two proteolytically activity fractions, HMW and LMW pyrolysin that were found to have identical NH_2-termini and were glycosylated to a similar extent. Additionally, autoincubation of the HMW pyrolysin resulted in a proteolytically active fraction with the size of LMW pyrolysin. Together, these data indicate that the LMW pyrolysin is generated from the HMW pyrolysin by the autoproteolytic removal of its COOH-terminal part.

Via reversed genetics the *pls* gene, coding for the pyrolysin of *P.furiosus*, was cloned and characterized. The deduced pyrolysin amino acid sequence consists of 1398 residues and is synthesized as prepro-enzyme, with the mature part of 1249 residues that showed the highest homology with eucaryal tripeptidyl-peptidases, that form a distinct subgroup of the subtilisin-like serine proteases (also referred to as subtilases). The NH_2-terminal catalytic domain (approximately 500 residues) showed considerable homology to subtilisin-like serine protease and contained a large insert of more than 150 residues between the aspartate and histidine active site residues. The COOH-terminal domain of pyrolysin together with the large insert in the catalytic domain contains almost all (29 of 32) of the possible *N*-glycosylation sites present in pyrolysin. Substrate specificity studies indicated that pyrolysin is a true endopeptidase with a rather broad substrate specificity and may autocatalytically remove its own propeptide.

The amino acid sequence homology in the catalytic domain of pyrolysin with other subtilases was sufficiently high to allow for homology modelling to gain insight in how the structure of the enzyme is stabilized. However, the identification of features as being important for protein stabilization based on only a single enzyme may lead to misinterpretations. Therefore, we screened a number of hyperthermophiles for the presence of subtilases using a PCR approach. Degenerated oligonucleotides were deduced based on the amino acid sequence of pyrolysin as well as on that surrounding the active site residues which are highly conserved among most of the subtilisin-like serine proteases. Chapter 7 describes the results of such a PCR approach that led to the identification of a DNA fragment in the genome of the extreme thermophilic archaeon

Thermococcus stetteri, which could encode a serine protease, designated stetterlysin. The deduced sequence of stetterlysin showed high homology with pyrolysin and contained a similar sized insert between the aspartate and histidine active site residues. The high homology between the two subtilases, stetterlysin and pyrolysin, and subtilases with an identified three dimensional structure allowed for homology modelling. Three-dimensional structure models for stetterlysin and pyrolysin were constructed based on the crystal structures of subtilases from mesophilic and thermophilic origin, respectively subtilisin BPN' and thermitase, and on the sequence alignment of the core residues of stetterlysin and pyrolysin with subtilisin and thermitase. The predicted model of subtilisin-S41 from psychrophilic origin was also included in the comparisons. The alignment and the predicted three-dimensional models were used to the analyze amino acid composition and structural features of the catalytic domain that could be related to thermostability. The higher thermostability of the subtilases, especially stetterlysin and pyrolysin, was found to be correlated with an increased number of residues involved in pairs and networks of charge-charge and aromatic-aromatic interactions. For stetterlysin and pyrolysin, most of the aromatic residues were located on the surface of the catalytic core and present in inserts within this domain, suggesting that for the overall structure of the proteases aromatic-aromatic interactions may have a even larger impact on the stability of the structure.

Analysis of the location of N-glycosylation sites in highly thermostable subtilases, including stetterlysin and pyrolysin, showed that most of these sites are located in surface loops. The modelling of the substrate binding region with known substrates was in good agreement with the observed broad substrate range for pyrolysin and the proposed autocatalytic activation (chapter 6).

The PCR approach with deduced oligonucleotides for subtilisin-like serine proteases also resulted in products with the expected size with DNA from the hyperthermophilic archaeon *Pyrodictium abyssi* and the hyperthermophilic bacterium *Fervidobacterium pennavorans*, that hybridized with a pyrolysin-derived probe and had the expected size. The PCR products were cloned and characterized and their deduced amino acid sequences showed significant homology with the catalytic domain of subtilases. However, unlike pyrolysin and stetterlysin, they did not contain the large inserts between the first two active site residues. Using this approach a *F.pennavorans* gene for a subtilisin-like serine has been isolated and is currently being characterized.

The hydrolases described in this thesis have been used as models to study different aspects of enzymes from hyperthermophiles, also referred to as thermozymes.

Industry has screened hyperthermophiles for enzymes that can be applied as industrial biocatalysts to replace less stable homologous in current processes or to initiate new processes. The industrial use of thermozymes is hampered by the cost-ineffective fermentation properties of the hyperthermophilic organisms, in which they reside. Therefore, functional overproduction of thermozymes in heterologous hosts may have considerable impact on the development of these enzymes as biocatalyst. This thesis describes the functional overproduction of number of thermozymes including a β-glucosidase an short-chain alcohol dehydrogenase and an endo-β-1,3-glucanase. Moreover, this work has contributed to the insight hydrolytic enzymes from hyperthermophilic archaea in the relation between their structure-function and structure-stability.

References:

Aagaard, C., Leviev, I., Aravalli, R.N., Forterre, P., Prieur, D., and Garrett, R.A. (1996) General vectors for archaeal hyperthermophiles: Strategies based on a mobile intron and a plasmid. *FEMS Microbiol. Rev.* **18**: 93-104.

Fischer, L., Bromann, R., Kengen, S.W.M., de Vos, W.M., and Wagner, F. (1996) Catalytic potency of β-glucosidase from the extremphile *Pyrococcus furiosus* in glucoconjugate synthesis. *Biotechnology* **14**: 88-93.

Jansen, A and Boon,F. (1997) Unpublished results.

Schulz, G.E. (1997) Unpublished results

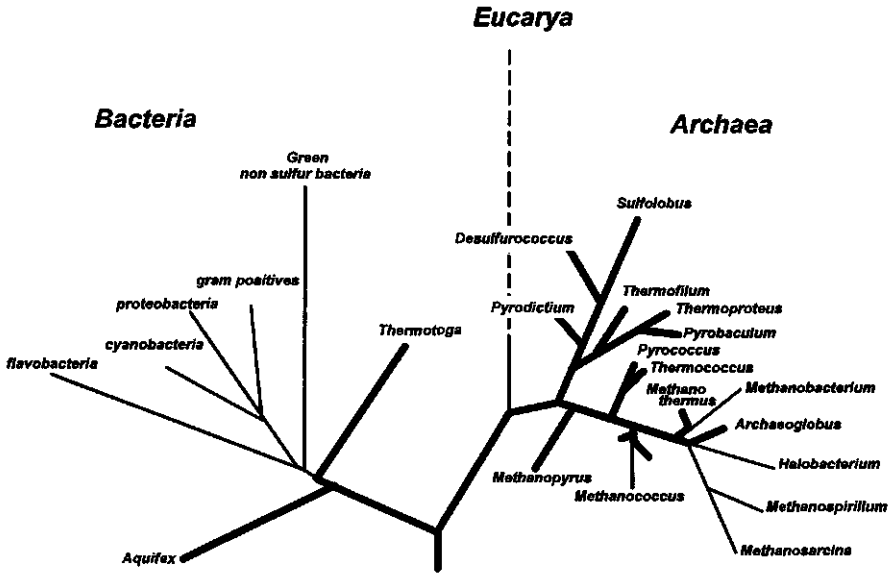
Sessitch, A., Wilson, K.J., Akkermans, A.D.L., and de Vos, W.M. (1996) Simultaneous detection of different *Rhizobium* strains marked with either the *Escherichia coli gus A* gen or the *Pyrococcus furiosus celB* gene. *Appl. Environm. Microbiol.* **62**: 4191-4194.

Trincone A, Pagnotta, E., Voorhorst, W.G.B., van der Oost, J., and de Vos, W.M. (1997) Stereochemistry of Glycosyl Transfer Using the β-glucosidase from the Hyperthermophilic Archaeon *Pyrococcus furiosus*. (Submitted to Carbohydrates).

Chapter 9

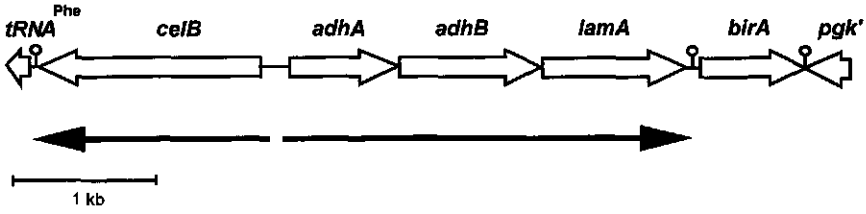
Samenvatting

Hyperthermofiele microorganismen worden gekenmerkt door groei rond het kookpunt van water en de meeste van hen behoren tot het domein van de *Archaea,* voorheen archaebacteriën. De *Archaea* vormen samen met de *Bacteria* en de *Eucarya* de drie domeinen van het leven op aarde (Fig. 1). De hoge temperatuur van de leefomgeving van deze opmerkelijke microorganismen maakt het noodzakelijk dat hun biomoleculen zoals DNA, RNA en eiwitten zijn aangepast aan deze en andere extreme condities. Zo zijn er eiwitten in hyperthermofielen gevonden die na 85 uur bij 100°C nog de helft van hun enzymatische activiteit bezitten. Naast hun thermostabiliteit zijn deze eiwitten vaak beter bestand tegen chemische denaturatie en minder gevoelig voor proteolyse dan homologen van mesofiele oorsprong. Vanwege deze bijzondere kenmerken is de interesse voor eiwitten van hyperthermofielen de laatste jaren sterk toegenomen vanuit zowel fundamenteel als ook biotechnologisch oogpunt. De fundamentele interesse is er ondermeer op gericht om inzicht krijgen in het proces van stabilisatie. De biotechnologische interesse komt voort uit de mogelijkheid om enzymen uit hyperthermofielen te gebruiken als biokatalysator en om de kennis van hun stabilisatie mechanismen toe te passen op biotechnologisch interessante enzymen uit mesofielen.



Figuur 1 De fylogenetische stamboom van het leven gebaseerd op 16S rRNA sequentie analyse met de drie domeinen aangegeven; *Bacteria, Eucarya* en *Archaea.* De dikke lijnen geven de hyperthermofiele microorganismen aan.

Het onderzoek beschreven in dit proefschrift is ondermeer gericht op het verkrijgen van inzicht in moleculaire mechanismen van adaptatie van eiwitten van hyperthermofiele archaea die leiden tot extreme stabiliteit bij hoge temperatuur. Tevens is getracht enzymen van hyperthermofielen te produceren in een heterologe gastheer met behoud van hun biochemische karakteristieken, waardoor de toepassings mogelijkheden van de enzymen uit hyperthermofielen vergroot worden. Om deze doelstellingen te realiseren werd de aandacht geconcentreerd op twee groepen van hydrolytische enzymen die een belangrijke plaats in nemen in het metabolisme van hyperthermofiele archaea: (i) glycosyl hydrolases die betrokken zijn bij de groei op polymere suikers waarvan de monomeren door middel van ß-verbindingen gekoppeld zijn, en (ii) serine proteases die een rol vervullen bij de groei op eiwitten en peptiden. De keuze voor hydrolytische enzymen komt ondermeer voort uit het feit dat homologen van mesofiele en thermofiele oorsprong goed gekarakteriseerd zijn op biochemisch, genetisch als ook op structureel niveau. Het organisme dat centraal heeft gestaan in dit proefschrift is de hyperthermofiele archaeon *Pyrococcus furiosus,* de meest uitvoerig bestudeerde vertegenwoordiger van de hyperthermofiele archaea. Regulatie van transcriptie in dit archaeon is bestudeerd aan de hand van twee divergente transcripten waarvan de produkten betrokken zijn bij de degradatie van ß-gekoppelde glucose polymeren. Het onderzoek dat in dit proefschrift beschreven wordt heeft met name betrekking op het *celB* locus van het hyperthermofiele archaeon *P.furiosus* en serine proteases van dit archaeon en van de verwante *Thermococcus stetteri.*



Figuur 2. Genetische en transcriptionele organisatie van het celB locus van het hyperthermofiele archaeaon *P.furiosus.*

In hoofdstuk 1 wordt een korte introductie gegeven in de verschillende aspecten van hyperthermofiele archaea. Dit omvat ondermeer het metabolisme van polymere substraten en de hydrolytische enzymen die hierbij betrokken zijn, de mechanismen die

een rol spelen bij de stabilisatie van eiwitten in hyperthermofielen, en een aantal opmerkelijke kenmerken van de moleculaire biologie van *Archaea.*

In hoofdstuk 2 wordt de isolatie en karakterisatie van het extreem stabiele β-glucosidase (halfwaarde tijd van 3 dagen bij 100°C) en het coderende *celB* gen van *P.furiosus* beschreven. Dit ß-glucosidase is essentieel voor *P.furiosus* bij de groei op de disaccharide cellobiose waar het de ß-1,4-glycoside binding tussen twee glucose residuen splitst. De transcriptionele organisatie van het *celB* gen werd geanalyseerd en suggereerde een aparte transcriptionele unit, hetgeen door Northern blot analyse ondersteund werd (zie hieronder) (Fig. 2). Het *celB* gen coderend voor het β-glucosidase werd via reversed genetics geisoleerd en de afgeleide aminozuur sequentie vertoonde de hoogste homologie met glycosyl hydrolases uit familie 1. Het *celB* van *P.furiosus* werd tot overexpressie gebracht in de mesofiele gastheer *Escherichia coli.* Dit resulteerde in een actief ß-glucosidase dat zowel kinetisch als structureel identiek bleek te zijn aan het in *P.furiosus* geproduceerde eiwit. De productie van dit hyperthermostabiel β-glucosidase (tot 20% van totale oplosbare eiwit fractie) in de mesofiele gastheer maakt de zuivering relatief eenvoudig door gebruik te maken van een hitte behandeling gevolgd door een enkele chromatografische stap. Het hoge productie niveau van het β-glucosidase in de genetisch goed toegankelijke gastheer *E.coli* maakte het mogelijk om door middel van 'protein engineering' inzicht te krijgen in de structuur-stabiliteit en -functie relatie. Met behulp van plaatsgerichte mutagenese kon worden aangetoond dat het mechanisme van katalyse voor het hyperthermostabiele ß-glucosidase van *P.furiosus* overeenkomt met dat wat gevonden wordt voor een homoloog uit de family 1 van glycosyl hydrolases dat optimaal actief is bij lage temperatuur.

Vanwege zijn opmerkelijke kenmerken is in de afgelopen jaren gekeken naar de mogelijkheden om het hyperthermostabiele ß-glucosidase van *P.furiosus* te gebruiken als biokatalysator. Door de brede substraatspecificiteit van het ß-glucosidase was het mogelijk om diverse produkten te synthetiseren in glucoconjugatie reacties in water en organische oplosmiddelen bij temperaturen van 75°C tot 95°C. Tevens is gebleken dat het ß-glucosidase van *P.furiosus* gebruikt kan worden voor de synthese van β-galacto-oligosaccharides uit lactose, die een belangrijke rol kunnen spelen in de voedingsmiddelen industrie vanwege hun potentiële functie als prebiotica (Jansen *et al.*, 1997). Daarnaast is het *celB* gen van *P.furiosus* ontwikkeld tot een genetische merker voor de studie van plant-bacterie interacties en competitie van *Rhizobium* stammen. De

toepassingsmogelijkheden van dit systeem heeft uiteindelijk geleid tot de ontwikkeling van de CelB Gene Marking Kit (FAO/IAEA).

Het hoge productie niveau van het pyrococcus β-glucosidase in *E.coli* en de eenvoudige zuiveringsprocedure maakte het mogelijk om voldoende gezuiverd eiwit te genereren zodat kristallisatie experimenten opgezet konden worden in samenwerking met de groep van Prof. G. Schulz (Universiteit van Freiburg). Dit heeft geresulteerd in kristallen die diffracteren met een resolutie van 3,5 Ångstrom (Schulz., 1997). Momenteel wordt geprobeerd om een drie-dimensionale structuur met hoge resolutie van het β-glucosidase van *P.furiosus* op te lossen, gebruik makend van de recent bepaalde structuur van het homologe LacS van *Sulfolobus solfataricus* in samenhang met plaatsgerichte mutagenese om een betere pakking te bewerkstelligen.

Het *celB* gen van *P.furiosus* is gelegen naast een in tegenover gestelde richting georienteerd gencluster coderend voor twee alcohol dehydrogenasen (*adhA, adhB*) en een endoglucanase (*lamA*) (Fig.2). In hoofdstuk 3 zijn de twee NADP(H)-afhankelijke alcohol dehydrogenases (ADHs) bestudeerd die elk tot een afzonderlijke groep behoren, de korte-keten ADH's (AdhA) en de ijzer-bevattende ADH's (AdhB). De primaire structuur van het AdhA van *P.furiosus* vertoont een hoge mate van homologie met korte-keten ADH's van bacteriële en eukaryele oorsprong. Dit kan als een aanwijzing gezien worden dat er al een korte-keten ADH aanwezig was in de laatste gemeenschappelijke voorouder van het leven, de progenoot. De primaire structuur van AdhB van *P.furiosus* vertoont de hoogste homologie met ijzer-bevattende ADHs, een relatief kleine groep die gekenmerkt wordt door ondermeer 4 geconserveerde histidine residuen. De *adhA* en *adhB* genen zijn tot expressie gebracht in *E.coli* hetgeen resulteerde in functionele eiwitten. Het AdhB, geproduceerd in *E.coli,* was in staat om methanol om te zetten in een NADP(H)-afhankelijke reactie, maar deze activiteit ging snel verloren waardoor er geen gedetailleerde karakterisatie plaats kon vinden. Het gezuiverde AdhA daarentegen vertoonde een stabiele NADP(H)-afhankelijke activiteit met een groot aantal primaire en secundaire alcoholen. Het substraat waar AdhA de hoogste activiteit mee vertoonde bestaat uit een koolstofketen van 5 residuen, met een voorkeur voor secundaire boven primaire alcoholen. De affiniteit van het *P.furiosus* AdhA bleek hoger voor aldehydes dan voor alcoholen en het enzym had een pH optimum voor deze reductie dat dicht bij de fysiologische pH ligt van *P.furiosus*. Het is dan ook het meest aannemelijk dat de fysiologische rol van AdhA bestaat uit de reductie van de aldehyde tot het corresponderende alcohol waarbij het de reductie-equivalenten verwijdert en cofactor regenereert. De lokatie van de *adhA* en *adhB* genen, in de

nabijheid van het *lamA* en het *celB* gen, maakt het verleidelijk te speculeren dat AdhA en AdhB beiden een rol vervullen in het metabolisme van ß-gekoppelde glucose polymeren.

In hoofdstuk 4 wordt de moleculaire karakterisatie van een endo-ß-1,3-glucanase en het hiervoor coderende *lamA* gen van *P.furiosus* beschreven. Het *lamA* gen ligt stroomafwaarts van de tandem *adhA-adhB* genen (Fig. 2). Dit endo-ß-1,3-glucanase van *P.furiosus* is in het domein van de *Archaea* de eerste vertegenwoordiger van de glycosyl hydrolase familie 16, die bestaat uit endo-ß-1,3- en endo-ß-1,3-1,4-glucanases. Voor de expressie in *E.coli* werd het *lamA* gen, zonder het DNA-fragment coderend voor de N-terminale leader sequentie, achter een T7 promoter geplaatst. Met dit expressiesysteem werd een functioneel en stabiel endo-ß-1,3-glucanase geproduceerd, dat 15 % van het totale oplosbare eiwitfractie van *E.coli* uitmaakte. Het gezuiverde enzym had de hoogste activiteit met ß-1,3-glucose polymeren, maar vertoonde ook activiteit met de ß-1,3-1,4-glucose polymeren lichenan en ß-glucan. Het endoglucanase vertoonde optimale activiteit bij pH 6-6.5 en een maximale activiteit bij 100-105°C, terwijl het bij 100°C de halfwaarde tijd had van 19 uur. Vergelijking van de aminozuur volgorde van glycosyl hydrolases van familie 16 geeft een onderverdeling te zien in een subgroep met endo-ß-1,3-1,4-glucanase activiteit en een tweede subgroep met voornamelijk endo-ß-1,3-glucanase activiteit. Deze laatste subgroep wordt gekarakteriseerd door een extra methionine residue in de voorspelde actieve site. Verwijdering van deze methionine in het endo-ß-1,3-glucanase van *P.furiosus* door middel van 'protein engineering' resulteerde niet in een veranderde substraat specificiteit, maar een verlaagde katalytische activiteit.

Het is gebleken dat *P.furiosus* in staat is om op het ß-1,3-gekoppelde glucose polymeer laminarine te groeien. Daarnaast bleek dat het endoglucanase *in vitro* laminarine te hydrolyseren tot oligomeren en glucose. Echter, door toevoeging van het ß-glucosidase van *P.furiosus* verliep de hydrolyse veel efficiënter. Deze complementaire activiteiten van het extracellulaire endoglucanase en het intracellulaire ß-glucosidase suggereren een essentiële rol voor deze enzymen bij groei op laminarine.

In hoofdstuk 5 wordt de regulatie van transcriptie beschreven van het *celB* gen en het divergent georiënteerde *adhA-adhB-lamA* gencluster in *P.furiosus*. Met behulp van Northern blot analyse kon worden aangetoond dat de *adhA-adhB-lamA* genen een 2.8-kb operon vormen, het *lamA* operon. Dit operon wordt stroomopwaarts geflankeerd door het *celB* gen dat een tegenovergestelde oriëntatie bezit en stroomafwaarts door het *birA* gen (Fig. 2). Het *celB* gen bleek overgeschreven te worden als monocistrone

messenger. De aanwezigheid van de enzymen, ß-glucosidase, alcohol dehydrogenase en endoglucanase, die gecodeerd worden door het *celB* gen en het *lamA* operon van *P.furiosus,* was in hoge mate afhankelijk van de aanwezige koolstofbron. De hoogste enzym niveau's werden gevonden in *P.furiosus* cellen die gegroeid waren op ß-gekoppelde glucose polymeren, zoals cellobiose en laminarine. Transcriptionele analyses van het *celB* gen en het divergent georiënteerde *lamA* operon gaven aan dat de controle plaats vindt op transcriptie niveau. Bovendien vindt er co-regulatie van de transcripten plaats en worden ze geïnduceerd door groei op ß-gekoppelde glucose polymeren. De transcriptie initiatie startplaatsen van het *celB* gen en het *lamA* operon lagen slechts 142 nucleotiden van elkaar verwijderd. De promotoren van beide genen vertoonden een hoge mate van conservering, waaronder de identieke archaele TATA-box. Analyse van transcriptie initiatie met behulp van een *in vitro* transcriptie systeem voor *P.furiosus* toonde aan dat transcriptie van het *celB* gen en het *lamA* operon op dezelfde plaats start als *in vivo*. Echter, de efficiëntie van transcriptie initiatie van zowel het *celB* gen als het *lamA* operon in dit systeem waren significant lager dan dat van het efficient overgeschreven *gdh* gen van *P.furiosus*. Dit suggereert dat een positieve regulator een bijdrage zou kunnen leveren aan de efficiëntie van transcriptie van de divergente transcripten van *celB* gen en het *lamA* operon in *P.furiosus*. Een mogelijke bindingsplaats voor een dergelijke regulator bevindt zich stroomopwaarts van de promotoren van het *celB* gen en het *lamA* operon. De resultaten van deze transcriptionele analyses maken het mogelijk om controle van transcriptie te bestuderen in het *in vitro* transcriptie systeem van *P.furiosus* en een regulator te identificeren die betrokken is bij de regulatie van de divergente genen van het *celB* locus. Daarnaast wordt een *in vivo* analyse van transcriptionele controle van het *celB* locus wordt wellicht haalbaar met de recente ontwikkelingen in het transformatie systeem voor *P.furiosus.*

Een combinatie van de huidige stand van zaken en inzichten met betrekking tot het metabolisme van β-gekoppelde glucose polymeren en de transcriptionele regulatie in het hyperthermofiele archaeon *P.furiosus* heeft geleid tot het volgende werkmodel (Fig. 3). De eerste stap is een depolymerisatie van grote polymere substraten die β-1,3-gekoppelde glucose verbindingen bevatte, zoals laminarine, tot kleinere fragmenten door het extracellulaire endo-β-1,3-glucanase (LamA). Na opname worden de kleine fragmenten intracellulair gehydrolyseerd door β-glucosidase activiteit van CelB. De complementaire activiteiten van deze twee glycosyl hydrolases kan inderdaad *in vitro* worden aangetoond. Ook de co-regulatie van transcriptie van de genen coderend voor

deze twee hydrolases impliceert een gezamelijke rol in de afbraak van specifieke glucose polymeren.

De laatste twee experimentele hoofdstukken van dit proefschrift beschrijven de karakterisatie en moleculaire modelling van serine proteasen die betrokken zijn bij de afbraak van caseïne en andere eiwit-achtige substraten. In hoofdstuk 6 wordt de zuivering en karakterisatie van het hyperthermostabiele serine protease pyrolysine en het coderende *pls* gen van *P.furiosus* beschreven. Het pyrolysine is een extracellulair protease dat geassocieerd is met de celmembraan van *P.furiosus*. De zuivering van pyrolysine werd voor een groot deel tot stand gebracht door een autoincubatie van de membraan fractie van *P.furiosus* cellen in 6 M ureum bij 95°C. Na de zuiveringsstappen bleven er twee proteolytisch actieve banden over, een hoog moleculaire en een laag moleculaire fractie, respectievelijk het HMW en LMW pyrolysine.

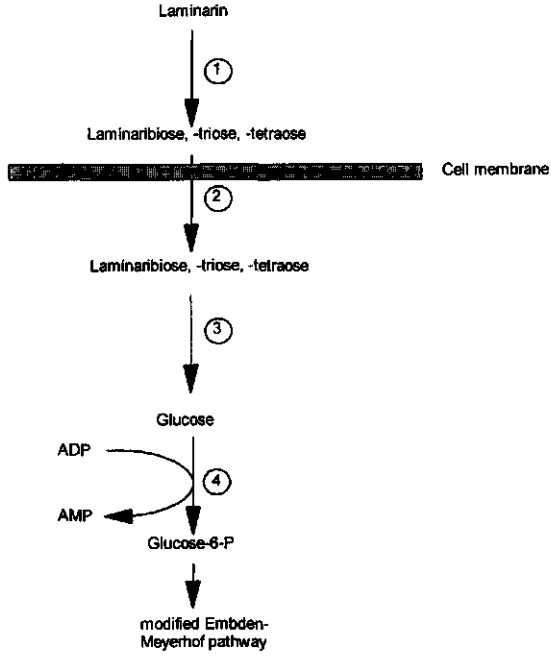


Figure 3. Model voor de groei op β-gekoppelde glucose polymeren zoals laminarin. 1, extracellulaire endo-β-1,3-glucanase (LamA); 2, onbekend opname systeem voor oligosacchariden; 3, β-glucosidase (CelB); 4, ADP-afhankelijk hexokinase.

De HMW en LMW pyrolysine fractie bleken identieke N-termini te hebben en vertoonden een vergelijkbare mate van glycosylering. Tevens resulteerde een autoincubatie van het HMW pyrolysine in een proteolytisch actieve fractie met dezelfde grootte als het LMW pyrolysine. De meest plausibele verklaring voor de resultaten is dat het LMW pyrolysine gevormd wordt door autoproteolyse van de C-terminus van het HMW pyrolysine.

Via reversed genetics werd het *pls* gen coderend voor pyrolysine van *P.furiosus* gekloneerd. De afgeleide pyrolysine sequentie omvat 1398 aminozuren en is opgebouwd uit een pre-pro domein dat gevolgd wordt door een matuur domein van 1249 aminozuren. Het laatste domein omvat ondermeer de active-site residuen karakteristiek voor subtilisine-achtige serine proteasen, ook wel subtilases genoemd, en vertoont een hoge mate van homologie met tripeptidyl-peptidases van eukaryele oorsprong, een subgroep van de subtilases. Het N-terminale katalytische domein (ongeveer 500 residuen) heeft tevens een hoge homologie met een andere subgroep van subtilases, hoewel het een insertie van meer dan 150 residuen bevat tussen de aspartaat en histidine active site residuen. Het C-terminale domein van pyrolysine bevat samen met het insert in het katalytische domein bijna alle mogelijke *N*-glycosylerings sites (29 van de 32) die in pyrolysine aanwezig zijn. Uit analyse van de substraatspecificiteit blijkt dat pyrolysine een endo-peptidase is met een relatief breed substraatbereik. Daarnaast verwijdert pyrolysine waarschijnlijk zijn eigen propeptide door middel van een autokatalytische reactie.

De hoge mate van homologie in de aminozuur volgorde in het katalytische domein van pyrolysine met dat van andere subtilases waar een drie-dimensionale structuur van bekend is maakt het mogelijk om op basis van deze bekende structuren een drie-dimensionale modelstructuur te bouwen voor pyrolysine. Deze zogenaamde homology modelling resulteerde in een model dat vervolgens geanalyseerd werd op structurele kenmerken die betrokken zouden kunnen zijn bij de stabilisatie van enzymstructuur.

De identificatie van factoren die belangrijk zijn voor eiwitstabilisatie op basis van één enzym kan echter tot verkeerde conclusies leiden. Daarom zijn een aantal hyperthermofielen onderzocht op de aanwezigheid van een subtilase gebruik makend van een specifieke amplificatie techniek, de polymerase chain reactie (PCR). In hoofdstuk 7 is beschreven hoe met behulp van deze PCR-techniek een gen geïsoleerd is uit de hyperthermofiele archaeon *Thermococcus stetteri* dat codeert voor stetterlysine, een serine protease waarvan de aminozuur volgorde een hoge mate van homologie

vertoont met die van pyrolysine. De hoge mate van overeenkomst tussen de twee serine proteasen uit *T.stetteri* en *P.furiosus*, en de subtilases waarvan de drie-dimensionale structuur zijn opgehelderd heeft homology modelling mogelijk gemaakt. Analyse van de modelstructuren heeft inzicht gegeven in factoren die een bijdrage zouden kunnen leveren aan de extreme stabiliteit van deze proteasen. Zo werd gevonden dat er een correlatie is tussen stabiliteit en de toename in het aantal zout-bruggen, zoals dat ook in ander eiwitten van hyperthermofiele oorsprong is gevonden. Tevens is er in de stabielere proteasen een significante toename gevonden in het aantal aromatische residuen, die betrokken kunnen zijn bij de vorming van interacties tussen aromatische residuen. Zowel in stetterlysine als in pyrolysine bleken de meeste aromatische residuen gelegen te zijn aan het oppervlak van de catalytische core en in het insert in dit domein. Dit suggereert dat voor de globale structuur van deze proteasen de interacties tussen aromatische residuen een belangrijke rol kunnen spelen.

Analyse van de locatie van *N*-glycosylerings sites in thermostabiele subtilases waaronder stetterlysin en pyrolysin, gaf aan dat de meeste van deze sites zijn gelokaliseerd in loops aan het oppervlak van de structuur. De modellering van de substraat bindings regio's met de geïdentificeerde substraten vertoonden een goede overeenkomst met het voorheen gevonden brede substraatbereik van pyrolysine en de veronderstelde autokatalytische activering (Hoofdstuk 6).

De PCR benadering die gebruikt werd voor de isolatie van subtilisine-achtige serine proteasen heeft geresulteerd in DNA fragmenten met de verwachte grootte voor het hyperthermofiele archeaon *Pyrodictium abyssi* en de hyperthermofiele bacterium *Fervidobacterium pennavorans*. Beide vertoonde hybridisatie met een van pyrolysine afgeleide probe. De PCR producten werden gekloneerd en gekarakteriseerd en de afgeleide aminozuur sequentie bleek een significante homologie te hebben met het catalytische domein van een aantal subtilases. Echter, in tegenstelling tot stetterlysine en pyrolysine, bevatten ze geen groot insert tussen de eerste twee active site residuen. Het PCR fragment van *F.pennavorans* werd vervolgens gebruikt voor de isolatie van chromosomale DNA fragmenten die het gehele gen bevatten en worden momenteel gekarakteriseerd.

De hydrolases die beschreven zijn in dit proefschrift zijn gebruikt als model om verschillende aspecten te bestuderen van enzymen van hyperthermofiele microorganismen, deze enzymen worden ook wel thermozymen genoemd. De industrie heeft in hyperthermofielen gezocht naar enzymen de gebruikt zouden kunnen worden als biokatalysator ter vervanging van minder stabiele homologe enzymen of om nieuwe

processen te initieren. Echter het gebruik van thermozymen in de industrie wordt bemoeilijkt door de dure fermentatie processen die nodig zijn om de hyperthermofiele microorganismen te kweken die de bewuste thermozymen bevatten. Functionele overproductie van thermozymen in heterologe gastheren zou daarom een sterke bijdrage kunnen leveren in de ontwikkeling van thermozymen als biokatalysator. Dit proefschrift beschrijft de functionele overproductie van een aantal thermozymen inclusief een β-glucosidase, een korte-keten alcohol dehydrogenase en een endo-β-1,3-glucanase. Tevens heeft het een bijdrage geleverd aan het inzicht in de structuur-stabiliteit en -structuur functie relatie van hydrolytische enzymen van hyperthermofiele archaea.

Referenties:

Aagaard, C., Leviev, I., Aravalli, R.N., Forterre, P., Prieur, D., and Garrett, R.A. (1996) General vectors for archaeal hyperthermophiles: Strategies based on a mobile intron and a plasmid. *FEMS Microbiol. Rev.* **18**: 93-104.

Fischer, L., Bromann, R., Kengen, S.W.M., de Vos, W.M., and Wagner, F. (1996) Catalytic potency of β-glucosidase from the extremphile *Pyrococcus furiosus* in glucoconjugate synthesis. *Biotechnology* **14**: 88-93.

Jansen, A and Boon,F. (1997) Unpublished results.

Schulz, G.E. (1997) Unpublished results

Sessitch, A., Wilson, K.J., Akkermans, A.D.L., and de Vos, W.M. (1996) Simultaneous detection of different *Rhizobium* strains marked with either the *Escherichia coli gus A* gen or the *Pyrococcus furiosus celB* gene. *Appl. Environm. Microbiol.* **62**: 4191-4194.

Trincone A, Pagnotta, E., Voorhorst, W.G.B., van der Oost, J., and de Vos, W.M. (1997) Stereochemistry of Glycosyl Transfer Using the β-glucosidase from the Hyperthermophilic Archaeon *Pyrococcus furiosus*. (Submitted to Carbohydrates).

Curriculum Vitae

Willebrordus Gerardus Bernardus (Wilfried) Voorhorst werd geboren op 7 augustus 1967 te Raalte. Na de basisschool is hij begonnen aan de HAVO (1979-1984) gevolgd door het VWO (1984-1986) op het Florens Radewijns College te Raalte. In 1986 is hij gestart met de studie Biologie aan de Landbouwuniversiteit Wageningen om na de propaedeuse de oriëntatie celbiologie te volgen. Het afstudeervak Industriële Microbiologie (Prof. J.A.M. de Bont) werd gevolgd door Moleculaire Biologie (Prof. dr. A. van Kammen). Zijn stages vonden plaats bij Shell Research Centre Sittingbourne, Engeland (Dr. A. Cornish) en de vakgroep Biochemie (Prof. dr. W.R. Hagen). Op 30 augustus 1992 sloot hij zijn universitaire studie af als ingenieur in de Landbouw- en Milieuwetenschappen. Op 12 augustus 1992 werd hij aangesteld als Assistent-In-Opleiding (AIO) bij de vakgroep Microbiologie van de Landbouwuniversiteit Wageningen. Binnen de werkgroep Bacteriële Genetica heeft hij gewerkt aan de moleculaire karakterisatie van hydrolytische enzymen van hyperthermofiele archaea, hetgeen uiteindelijk in dit proefschrift heeft geresulteerd.

List of publications:

Sander A. Peters, Wilfried G.B. Voorhorst, Jan Wery, Joan Wellink, and Albert van Kammen (1992) A Regulatory Role for the 32K Protein in Proteolytic Processing of Cowpea Mosaic Virus Polyproteins. **Virology**, 191, 81-89

Sander A. Peters, Wilfried G.B. Voorhorst, Joan Wellink, and Albert van Kammen (1992) Processing of VPg-Containing Polyproteins Encoded by the B-RNA from Cowpea Mosaic Virus. **Virology**, 191, 90-97

André de Haan, Mark R. Smith, Wilfried G.B. Voorhorst, and Jan A.M. de Bont (1993) Co-factor Regeneration in the Production of 1,2-epoxypropane by *Mycobacterium* strain E3: The Role of Storage Material. **Journal of General Microbiology**, 139, 3017-3022

Wilfred R. Hagen, Marc F.J.M. Verhagen, Antinio J. Pierik, Ronnie B.G. Wolbert, Leonard F. Mallée, Wilfried G.B. Voorhorst (1993) Axial Coordination and Reduction Potentials of the 16 Hemes in High-Molecular-Weight Cytochrome c from *Desulfovibrio vulgaris*. **Journal of Inorganic Biochemistry**, 51, 28

Marc F.J.M. Verhagen, Wilfried G.B. Voorhorst, Joost A. Kolkman, Ronnie B.G. Wolbert, and Wilfred R. Hagen (1993) On the Two Iron Centers of Desulfoferrodoxin. **FEBS Letters**, 336, 13-18

Marc F.J.M. Verhagen, Antonio J. Pierik, Ronnie B.G. Wolbert, Leonard F. Mallée, Wilfried G.B. Voorhorst, and Wilfred R. Hagen (1994) Axial Coordination and Reduction Potentials of the Sixteen Hemes in High-Molecular-Mass Cytochrome *c* from *Desulfovibrio vulgaris* (Hildenborough). **European Journal of Biochemistry**, 225, 311-319

Rik I.L. Eggen, Ans C.M. Geerling, Wilfried G.B. Voorhorst, Remco Kort, and Willem M. de Vos (1994) Molecular and Comparative Analysis of the Hyperthermostable *Pyrococcus furiosus* Glutamate Dehydrogenase and Its Gene. **Biocatalysis**, 11, 131-141

Wilfried G.B. Voorhorst, Rik I.L. Eggen, Evert J. Luesink, and Willem M. de Vos (1995) Characterization of the *celB* Gene Coding for ß-Glucosidase from the Hyperthermophilic Archaeon *Pyrococcus furiosus* and Its Expression and Site-Directed Mutation in *Escherichia coli*. **Journal of Bacteriology**, 177, 7105-7111

Wilfried G.B. Voorhorst, Rik I.L. Eggen, Ans C.M. Geerling, Christ Platteeuw, Roland J. Siezen, and Willem M. de Vos (1996) Isolation and Characterization of the Hyperthermostable Serine Protease, Pyrolysin, and Its Gene from the Hyperthermophilic Archaeon *Pyrococcus furiosus*. **Journal of Biological Chemistry**, 271 (34), 20426

Wilfried G.B. Voorhorst, Angela Warner, Willem M. de Vos, and Roland J. Siezen (1997) Homology Modelling of Two Subtilisin-like Serine Proteases from Hyperthermophilic Archaea *Pyrococcus furiosus* and *Thermococcus stetteri*. **Protein Engineering**, 10 (8),905-914

Yannick Gueguen, Wilfried G.B. Voorhorst, John van der Oost, and Willem M. de Vos (1997) Cloning and Sequencing of the *lamA* gene from the Hyperthermophilic Archaeon *Pyrococcus furiosus* Coding for an Endo-ß-1,3-glucanase and Its Expression and Site-directed Mutation in *Escherichia coli* **Journal of Biological Chemistry**, 272, 31258-64

Wilfried G.B. Voorhorst, Vincent Wittenhorst, John van der Oost, and Willem M. de Vos (1997) Genetic and Biochemical Characterization of a Short-chain and an Iron-containing Alcohol Dehydrogenase from the Hyperthermophilic Archeaon *Pyrococcus furiosus*.(submitted for publication)

Wilfried G.B. Voorhorst, Yannick Gueguen, Gerti Schut, Isabell Dahlke, Michael Thomm, John van der Oost, and Willem M. de Vos (1997) Transcriptional Regulation in the Hyperthermophilic Archaeon *Pyrococcus furiosus*: Coordinated Expression of Divergently Transcribed Genes in Response to ß-linked Glucose Polymers (submitted for publication)

Antionio Trincone, Edoardo Pagnotta, Wilfried G.B. Voorhorst, John van der Oost, and Willem M. de Vos (1997) Stereochemistry of Glycosyl Transfer Using the β-glucosidase from the Hyperthermophilic Archaeon *Pyrococcus furiosus*. (Submitted to Carbohydrates)

Willem M. de Vos, Servé Kengen, Wilfried G.B. Voorhorst, and John van der Oost (1998) Sugar Utilization and Its Control in Hyperthermophiles. Extremophiles (In press)